MODELLING PHOSPHINE SPECTRA FOR THE ATMOSPHERIC CHARACTERIZATION OF COOL STARS AND EXOPLANETS

Clara Sousa-Silva

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of the

University College London.

Department of Physics & Astronomy

University College London

August 3, 2015

to my parents

Associated Publications

- ExoMol line lists VII: The rotation-vibration spectrum of phosphine up to 1500 K

  C. Sousa-Silva, Ahmed F. Al-Refaie, S. N. Yurchenko and J. Tennyson

  Monthly Notices of the Royal Astronomical Society, 446: 2337-2347, 2015.

- High Temperature Partition Functions and Thermodynamic Data for Ammonia and Phosphine

  C. Sousa-Silva, N. Hesketh, S. N. Yurchenko, C. Hill and J. Tennyson

  Journal of Quantitive Spectroscopy and Radiative Transfer, 142: 66-74, 2014.

- A Computed Room Temperature Line List for Phosphine

  C. Sousa-Silva, S. N. Yurchenko and J. Tennyson

  Journal of Molecular Spectroscopy, 288: 28-37, 2013.

Abstract

Understanding the molecular composition of astrophysical atmospheres depends on the interpretation of their spectra, which is possible through a combination of observational data analysis and computational modelling. However, this process is difficult to perform as the observed spectra have an extremely rich structure and their opacity is dominated by multiple and complex molecules. These issues can only be addressed through a sophisticated understanding of molecular behaviour, together with fundamental data for all the species that contribute significantly to an atmosphere's opacity.

The present work describes the motivation and requirements for the creation of molecular line lists, and their applications. It specifically focuses on phosphine, PH_3, due to its relevance in both earthly and astronomical studies, in particular cool stars and exoplanets. A complete line list for phosphine is presented, having been generated by employing *ab initio* quantum mechanical methods, performing an empirical refinement based on experimental spectroscopic data and harnessing high performance computing.

A full account of how the spectral data were obtained, from the creation of the requisite surfaces to the production of cross-sections, can be found in these pages. An analysis of the partition function and associated thermodynamic properties of both phosphine and ammonia, NH_3, is also done. Additionally, ongoing work from a preliminary investigation into the use of selected symmetries to produce fast and inexpensive low resolution spectra is described, and results from an introductory study of the tunnelling motion of phosphine are presented.

The connections between past, present and future work are discussed, as well as their applications to the fields of astronomy and molecular physics.

Acknowledgements

The author would like to thank her excellent supervisors Sergey N. Yurchenko and Jonathan Tennyson, for their innumerable contributions and assistance; co-authors Christian Hill and Nicholas Hesketh for their involvement in the Thermodynamic Properties paper; co-author Ahmed Al-Refaie for his incredible algorithm without which the SAlTY phosphine line list might still be incomplete and who, together with Duncan Little, made UCL a home; my patient and ever encouraging housemates Hannah Lees, Sally Hewett and Helen Parks; Jane Clossick, without whom a chapter that did not deserve to exist would still be under construction and who, together with Antonio S. Silva and Dominic Bristow, painstakingly read through most of the thesis, offering constructive and supportive feedback; My dad, Antonio Silva and friends Minnie Ashdown, Peter Day, Anna McInerney, Rosie Bell, Colin O'Sullivan and Charles Leahy for the support, advice and respite they provided.

This work was supported by the ERC Advanced Investigator Project 267219. The research made use of the DiRAC@Darwin, DiRAC@COSMO and the EMERALD HPC clusters. DiRAC is the UK HPC facility for particle physics, astrophysics and cosmology and is supported by STFC and BIS. The EMERALD High Performance Computing facility provided via the Centre for Innovation (CfI). The CfI is formed from the Universities of Bristol, Oxford, Southampton and UCL in partnership with STFC Rutherford Appleton Laboratory.

Contents

List of Figures

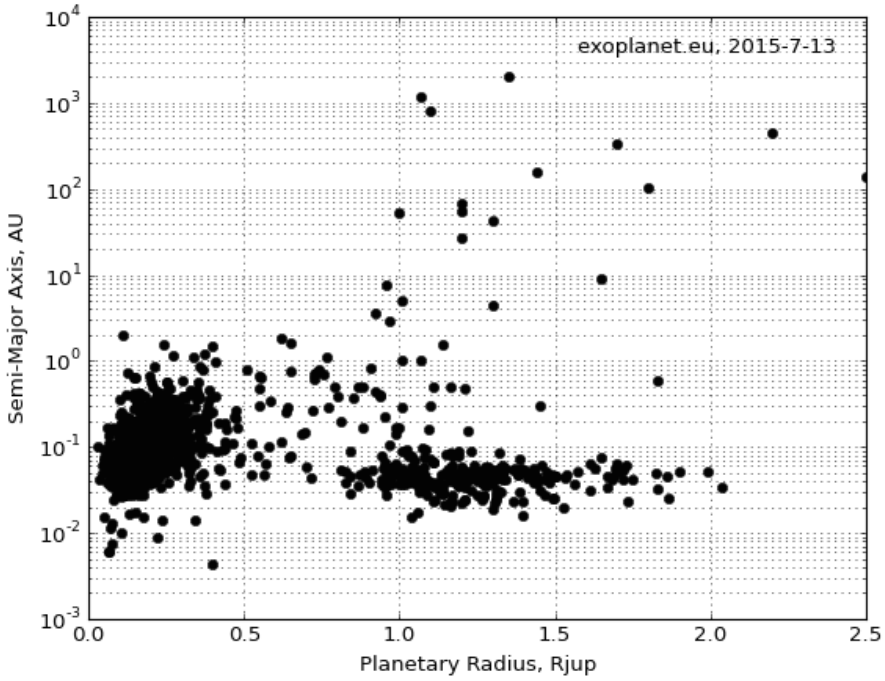
List of Tables

Chapter 1

Introduction

There are over one hundred thousand million stars in the Milky Way, the majority of which are expected to have planetary systems [52]. This exoplanet ubiquity is due to their development as a collateral effect of stellar formation; when stars are forming by accumulation of material from their protoplanetary disk, some of that material will aggregate into substellar bodies, some of which will become planets. This planetary pervasiveness has led to almost 2000 detections over the past twenty years (see figure 1.1), with new detections routinely being made.

Figure 1.1: Exoplanet discoveries as of the 9th April 2015, plotted with semi-major axis (AU, log scale) vs planetary radius in units of Jupiter radii. One AU corresponds to the distance from the Earth to the Sun. Figure taken from *exoplanet.eu*.



Over the past twenty years, discoveries of stars, substellar objects and planets have demonstrated the enormous variety of astrophysical bodies in the Universe. Exoplanets are much more diverse than it was first predicted a decade ago, when the field was in its infancy. Stars, brown dwarfs, gas planets and rocky planets can all be roughly considered to be on a continuous planet-star spectrum, with bodies continuously being found that question the boundaries of these previously delineated classifications. Furthermore, the diversity in temperature, densities, composition and atmospheric parameters of these bodies is extremely large, and is expected to keep on growing. Examples of unusual planetary possibilities include lava planets [76], water worlds [149] and diamond planets [100]. Now that it is known that innumerable worlds exist with a vast wealth of characteristics, attention is focusing on understanding and characterising these bodies.

The systematic monitoring of stars in the neighbourhood of the Solar System has significantly increased our understanding of stars and exoplanets. Detection and analysis through transit studies, radial velocities, photometry, microlensing, and direct imaging lead to rich data sets that yield more information about these bodies every day. Additionally, knowledge of stellar and planetary formation, models of interior structure and planetary dynamics have all contributed to these studies. However, without understanding the atmospheres of stars and exoplanets, it is impossible to fully characterise them.

Building on previously released ExoMol line lists for polyatomic molecules, this thesis explores the work undertaken during the creation of a comprehensive high temperature line list for phosphine. The remainder of the introductory chapter will briefly describe the processes behind the extraction of molecular profiles from astrophysical atmospheres and detail the importance of phosphine in such environments. It focusses on atmospheric spectra and offers an overview of the processes and calculations involved in characterising an atmosphere. This chapter sets up key preliminary information for the subsequent chapters and it includes an overview of the steps between starlight, observational spectra and the inference of the molecular composition of an atmosphere.

Chapter 2 is the key theoretical background chapter, covering line list theory. It offers an overview of the processes behind creating a molecular line list from first principles, and the methods for presenting it in a format that can be used to interpret observational spectra.

Chapters 3, 4 and 5 are based on the associated publications listed on page 3. The first of these papers (Chapter 3) concerns phosphine and the specific requirements and obstacles sur-

rounding its simulation. It describes in detail the calculation of energy levels, the transitions between them and corresponding line strengths, which led to the creation of a preliminary, cool line list for phosphine, called SYT (**S**ousa-Silva, **Y**urchenko and **T**ennyson). The second paper (Chapter 4) is a description of the calculation of the thermodynamic properties of phosphine and ammonia, by explicit summation of previously computed energy levels. The third paper (Chapter 5) represents the culmination of the work presented in this thesis, and concerns the computation of the final, hot phosphine line list, called SAlTY (**S**ousa-Silva, **Al**-Refaie, **T**ennyson, **Y**urchenko). It considers improvements to the preliminary line list and the associated difficulties as well as a discussion of the line list formats. Each of these chapters is organised with a description of its methods and a discussion of the findings.

The final chapter is the conclusion, bringing together the discussions from each of the three core chapters. It presents a consideration on the future atmospheric modelling of phosphine, and introduces two ongoing projects that follow on from the work presented in this thesis, Phantom Symmetries and Phosphine Tunnelling.

1.1 Atmospheric Analysis

The most promising route towards characterising exoplanets is the spectroscopic analysis of their atmospheres and that of their host stars, by either transit spectroscopy or direct imaging. Different molecules preferentially absorb and emit light at different but very specific frequencies; studying the selective absorption and emission from a light source can provide insight into the molecules of the environments the light travelled through. This process is further described in Chapter 2.

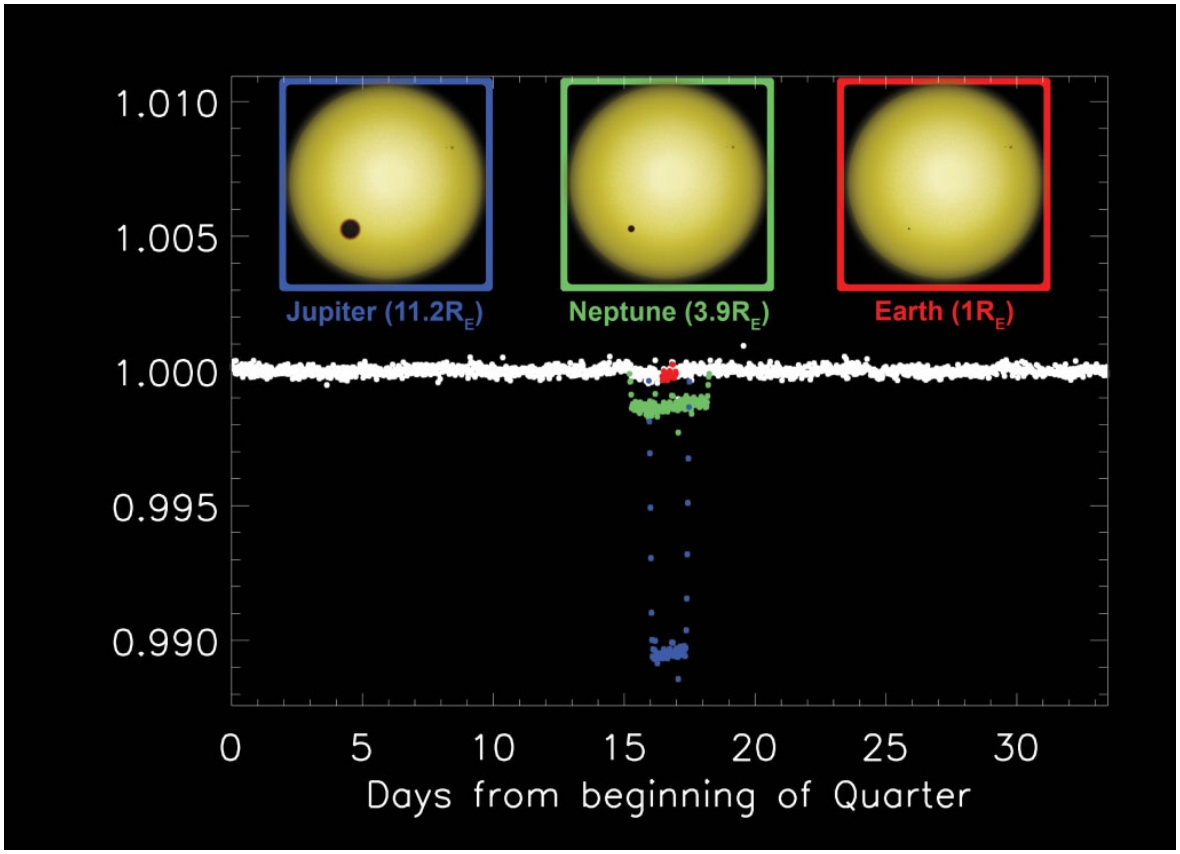
Molecules can be found in many astronomical environments other than cool stars and exoplanets, such as comets, interstellar clouds, planetary nebulae and the tori formed around active galactic nuclei [160]. The most important tool in the detection and characterisation of these molecules is spectroscopy. Unlike black bodies, whose spectra are dependent only on temperature, it is possible to gain insight into the various molecular properties of astrophysical bodies by spectroscopic analysis of their light. There are various processes for obtaining this light, with the most prevalent ones described below. Given the right data, it is possible to study that light and investigate the presence, abundance and environment of any molecule creating an observable spectral signature.

1.1.1 Transit Spectroscopy

From the perspective of the Earth, when planets pass in front of their host star a fraction of the star's light is blocked out, and a decrease in flux is observed; this is called a primary transit. Over the course of the whole transit, a time dependent light curve is created. The depth of this curve is dependent on a variety of parameters, such as the relative sizes of the star and the planet, the orbital dynamics of the system and, most importantly in the context of this work, the composition of the planetary atmosphere (if it is present).

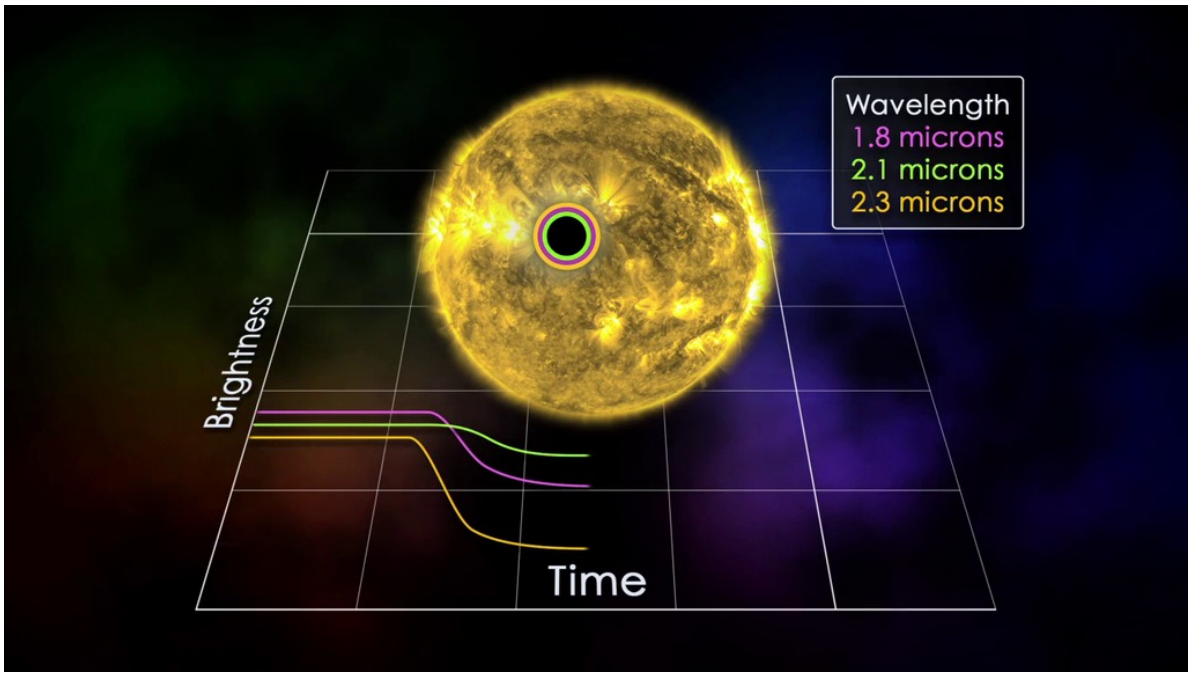
However, the relative size between exoplanets and their host star makes it difficult to retrieve accurate information on these parameters. A Jupiter sized planet would at most block $\approx 1\%$ of its host star's light; for the Earth this decreases to $\approx 0.01\%$. This is due to their relative radii, which roughly follow a ratio of $1 : 10 : 100$ for the Earth : Jupiter : Sun. The light depth created by an Earth-sized transit would be barely perceptible, as presented in Figure 1.2, which demonstrates the effect of a Jupiter, Neptune and Earth sized planet would have on a Sun-like star's light curve when transiting across it.

Figure 1.2: Light curves for a Jupiter, Neptune and Earth-like transit. Figure taken from *zooniverse.org*.



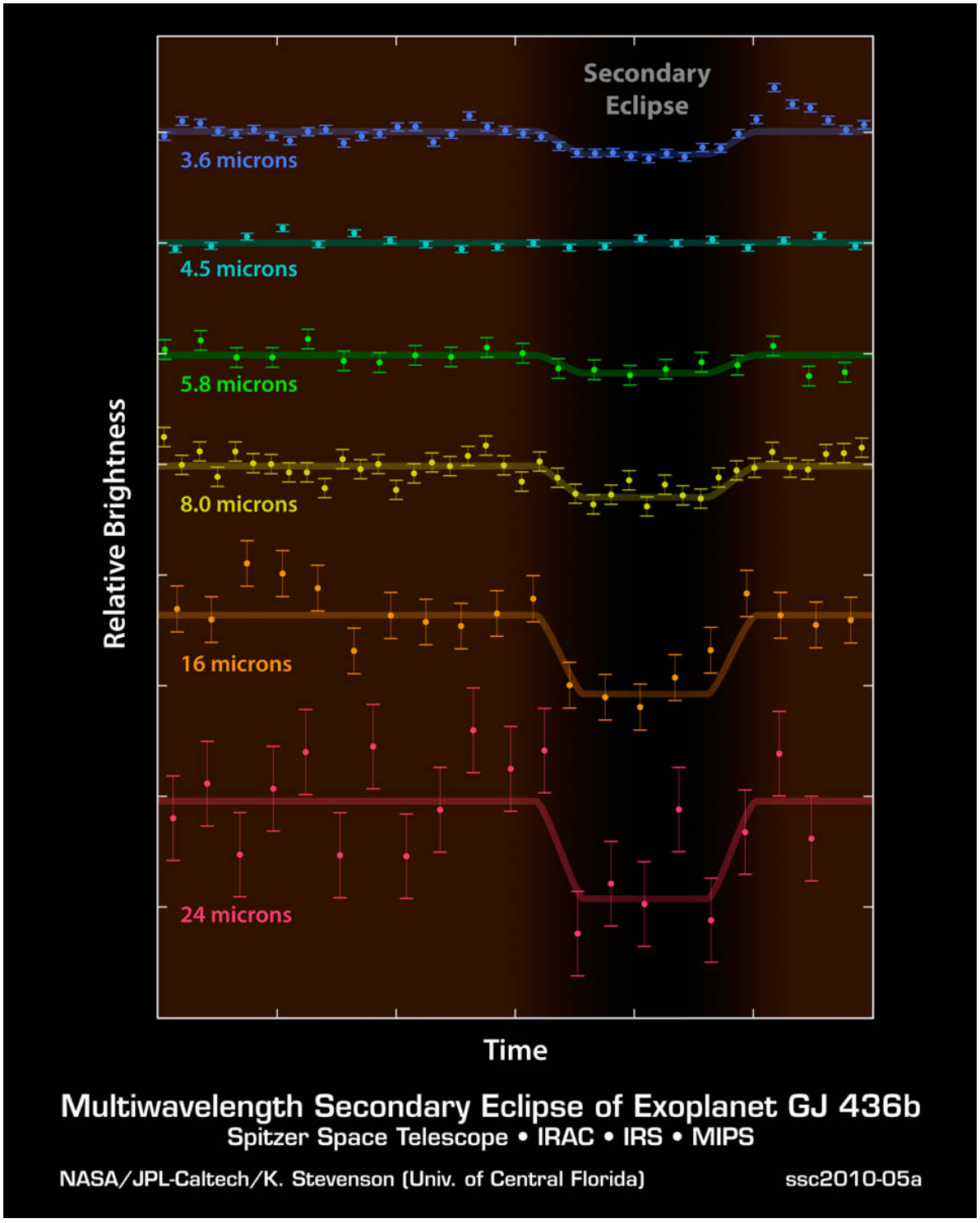
If a planet transiting a star has no atmosphere, its transit depth is the same at all wavelengths, since it is only a representation of the physical obstruction of starlight caused by the radius of the planet. However, if the planet does have an atmosphere, atoms and molecules in it will absorb some of the light from the star, and the apparent radius of the planet will change with wavelength, as the molecules in the atmosphere preferentially absorb different parts of the light spectrum. Figure 1.3 demonstrates how the atmosphere of a planet may cause the depth of the transit to vary with wavelength as the planets crosses in front of its host star. The relative depth of the light curve at these wavelengths can then be associated to the absorption fingerprint of specific molecular and atomic species.

Figure 1.3: Light curve dependence on wavelength for a transiting planet with an atmosphere. Figure credited to NASA's Goddard Space Flight Center.



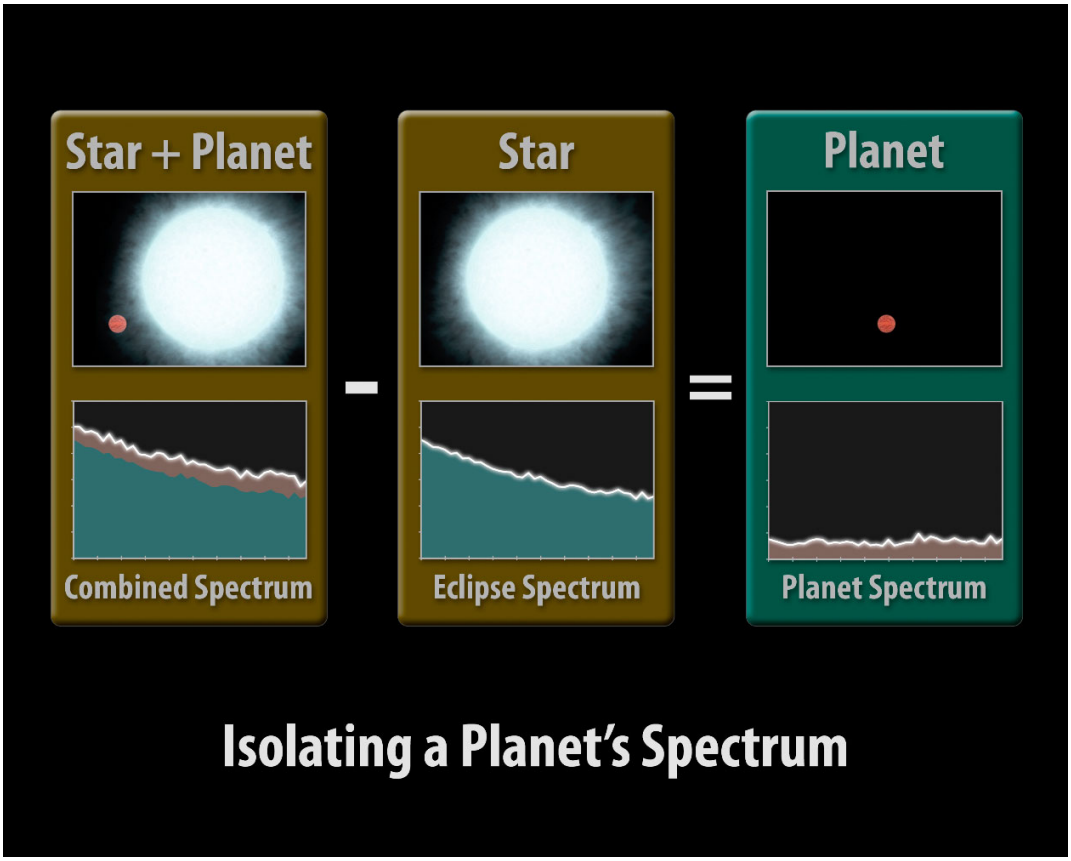
When the planet moves entirely behind the star, any contribution to the spectra from the planet is removed from the light curve; this is known as a secondary transit. Figure 1.4 shows the light curve variation with wavelength during the secondary transit of the hot Neptune GJ 436b. Secondary transits cause a smaller depth in the light curve than primary transits, as the thermal radiation from the planet only constitutes a small fraction of the total light of the star/planet system.

Figure 1.4: Light curve dependence on wavelength for the secondary transit of the hot Neptune GJ 436b. Figure credited to NASA/JPL-Caltech/K. Stevenson, data from Spitzer's IRAC, IRS and MIPS [178].



During the secondary transit, the observed stellar spectrum is unadulterated by the planet's atmosphere. Thus, a comparative study of the spectra from the primary and secondary transits can provide an absorption spectra for which the planet's atmosphere is solely responsible (see, as an illustrative example, Figure 1.5).

Figure 1.5: Simulation of the light spectra for the Hot Jupiter HD 209458b using primary and secondary transit data. Figure taken from *NASA*.



Retrieving the contribution to the spectra for which the exoplanet atmosphere is exclusively responsible is not a straightforward process. It is vital that information on the host star is available, as that is the only way the stellar spectra can be completely extracted from the combined planet-star spectra. The contribution to the spectra from the star alone depends on stellar activity and solar spots, which can be highly complex to model and require sophisticated interpretation. In addition, most models rely on solar type abundances, which often do not correspond to the metallicities of the astrophysical bodies being studied. Finally, the atmospheric dynamics of the planet itself will also affect the resulting spectra; for example, cloud formation will have a significant impact on the light absorption at different altitudes[105].

1.1.2 Direct Imaging

It is sometimes possible to obtain atmospheric spectra directly by imaging a planet, rather than just considering it as an interference in its host star spectra. Direct imaging provides spectra for stars and substellar objects, but it is also possible to use this method to observe planetary companions. However, this is difficult due to the flux ratio between exoplanets and their host star, as most planets either produce very little to no light of their own, or reflect only a minute fraction of their host star's flux. For example, for a Jupiter-like planet in a solar type system, the flux ratio is $\approx 10^{-9}$ at maximum elongation, with an angular separation of 0.5 arc sec at 10pc. For an Earth-like planet, the flux ratio decreases to $\approx 10^{-10}$ [131].

There are still very few of these directly imaged atmospheric spectral retrievals, although those that exist have already provided a wealth of knowledge about their targets (see, for example, [125] and [154]). Retrievals are also difficult to achieve with a good signal to noise ratio (SNR), and sufficient wavelength coverage for a full analysis of molecular composition. This is a problem present in other detection methods, but direct imaging is particularly affected by stellar activity, making accurate retrievals additionally difficult.

In general, obtaining clear atmospheric spectra for exoplanets requires overcoming a variety of obstacles, including competing flux sources, incomplete knowledge of the formation, dynamics and composition of the system, as well as uncertainties introduced by stellar variability, like starspots and other stellar processes. Ultimately, if these issues are resolved, it is the quality of the available data on molecular fingerprints that will be the limiting factor for the correct interpretation of spectral features from astrophysical atmospheres.

1.1.3 Molecular Fingerprints

Cooler astronomical bodies than our Sun have a large and varied molecular presence in their outer atmospheres. Alongside atomic spectra, hundreds of molecules contribute towards the atmospheric spectral signature of these cool stars and exoplanets. This molecular wealth leads to rich, complex spectra with a multitude of features which may be broadened by the effects of pressure and temperature [8].

Even if a planet's atmospheric spectrum were to be adequately extracted, it is usually far too sophisticated to be interpreted in isolation. To be able to interpret observed spectra, atmospheric models need to be simulated that can be compared to the retrieved astronomical

data. If a reasonable set of parameters can be found that create an atmospheric model fitting the observed data points, constraints can be put on the composition and structure of a planet's atmosphere. This process and its variants are described in detail in multiple sources (e.g. [6, 72, 151, 101]).

Planetary atmospheres are analysed using both observational data from the light of the host star (or the planet directly) and sophisticated models to fit this data; in either case, understanding what chemical composition could result in a particular transmission spectra requires access to detailed information about each of its atomic and molecular components. Without knowing the exact spectral fingerprint of each molecule, an accurate description of the composition of the target is impossible. Independently of the sources of uncertainty inherent in atmospheric observations and associated models, the availability of complete and accurate fundamental molecular descriptions is crucial for the correct characterisation of an atmosphere.

The causes for inconsistencies between models and observations can be varied, but the source of uncertainty brought about by poor and incomplete molecular opacity databases cannot be underestimated [110]. For example, many astronomical spectra have unidentified features which correspond to molecules that cannot yet be detected due to the lack of molecular data available.

Apart from a few exceptions (e.g. H_2O, CO_2, CO, NO and OH [141]), knowledge of the spectral profiles for the molecules of interest in astrophysical atmospheric studies is poor; much is incomplete, inaccurate or unreliable, and, for many key species, non-existent. The production of reliable, complete molecular spectra is the driving force behind the ExoMol project [165] and the associated work present here.

1.2 The ExoMol Project

The ExoMol project aims to create a reliable molecular database (*www.exomol.com* [165]), which seeks to provide systematically comprehensive line lists for the molecules relevant to the atmospheric studies of planets and stars.

Each molecule is simulated using a combination of *ab initio* and empirically-tuned quantum mechanical methods. It is then presented as a rotation-vibration and rotation-vibration-electronic line list, with a catalogue of transition frequencies and their associated probabilities. Although some molecules are more challenging to simulate than others, this can be an extremely computationally demanding process so ExoMol makes use of multiple high performance computing centres. Though the primary concern of ExoMol is the characterisation of cool stars and exoplanets, the applications of molecular data reach beyond this.

Molecular line lists have been used in a variety of different fields including, among others, modelling the Earth's atmosphere [34], circumstellar envelopes, comets [13, 142], remote detection of forest fires, imaging in gas engines [90], analysing the temperature profile in flames, engines and exhaust gases, studying the cosmic ray background [147], and many others.

Many molecular spectra have been previously measured and simulated. Some experimental databases, such as HITRAN [139], GEISA [75], JPL [133] and CDMS [112], contain many thousands of transitions for each molecule in their catalogue, though they only include a few species. However, these databases simply do not contain enough information, as molecular spectra can become extremely rich, particularly at high temperatures, with billions of transitions contributing to the shape of the spectrum of a molecule. A few other databases exist, but are even more inaccurate or incomplete than those mentioned above. Additionally, experimental databases usually have limited spectral range and are only appropriate for modelling environments with temperatures below 296 K, as it is challenging, expensive and often dangerous to measure molecular spectra at high temperatures. HITEMP [141], HITRAN's high temperature database, does contain data appropriate for modelling molecular spectra at elevated temperatures. However, it does not include PH_3, consisting only of five species (H_2O, CO_2, CO, NO and OH).

The methods and obstacles involved in simulating a molecule are discussed throughout this thesis, but it is a difficult and expensive process. As such, the ExoMol team is effectively divided into molecules, with each member spending multiple years becoming extremely fa-

miliar with their assigned molecule. For the past few years, it has been the author's personal responsibility to consider phosphine.

1.3 The Phosphine Molecule

Phosphorus is the one of the most abundant chemically reactive volatile elements in a solar type system (with S, after H, C, N and O, collectively known as CHNOPS species). Although phosphorus has a considerably smaller cosmic abundance than H, C, O or N, it is predicted to have an important role in atmospheric chemistry and dynamics throughout the galaxy. Phosphorus is not particularly common in the universe but it is ubiquitous and is important for most essential biochemical functions on Earth. Due to its role as a biogenic particle, phosphorous has been put forward as a potential marker in the search for extinct or existing life in other planets [99].

Phosphorous speciation is substantially more complex than that of the other CHNOPS species, due to its disequilibrium chemistry [173]. A large fraction of the existing phosphorus in various astronomical environments is expected to be found in the form of phosphine, or PH_3 [1, 173]. Different phosphorus species are dominant at particular temperature and pressure combinations, and phosphine abundances vary greatly with the local environment.

Throughout, PH_3 and phosphine refer to the main isotopologue $^{31}PH_3$. Phosphorus has 23 isotopes, but ^{31}P is the only non-synthetic, stable isotope of phosphorus, with 16 neutrons. Phosphorus is therefore considered a monoisotopic element.

In nature, phosphine is produced during the anaerobic decomposition of carbon based organic matter [57, 56, 116], but it rapidly breaks down in the Earth's atmosphere, as it has a half-life in air of less than a day [113]. Its biogenesis is not fully understood but it is a natural and crucial part of the phosphorous cycle on Earth; it plays an important role in the Earth's hydrosphere and has been found in the Antarctic biosphere [196] and in the lower and upper troposphere of the Earth [34, 42].

Phosphine gas is colourless and odourless, but it is corrosive, spontaneously flammable and extremely toxic [180]. Since it can be produced by organic sources, it has been suggested as a presumably humorous, albeit chemically sound, explanation for fire breathing dragons [56]. Phosphine is used in agriculture, for pesticides and as a fumigator, and in a few other industries, such as metallurgy, polymers and semi-conductors [87]. It is also created as a by-

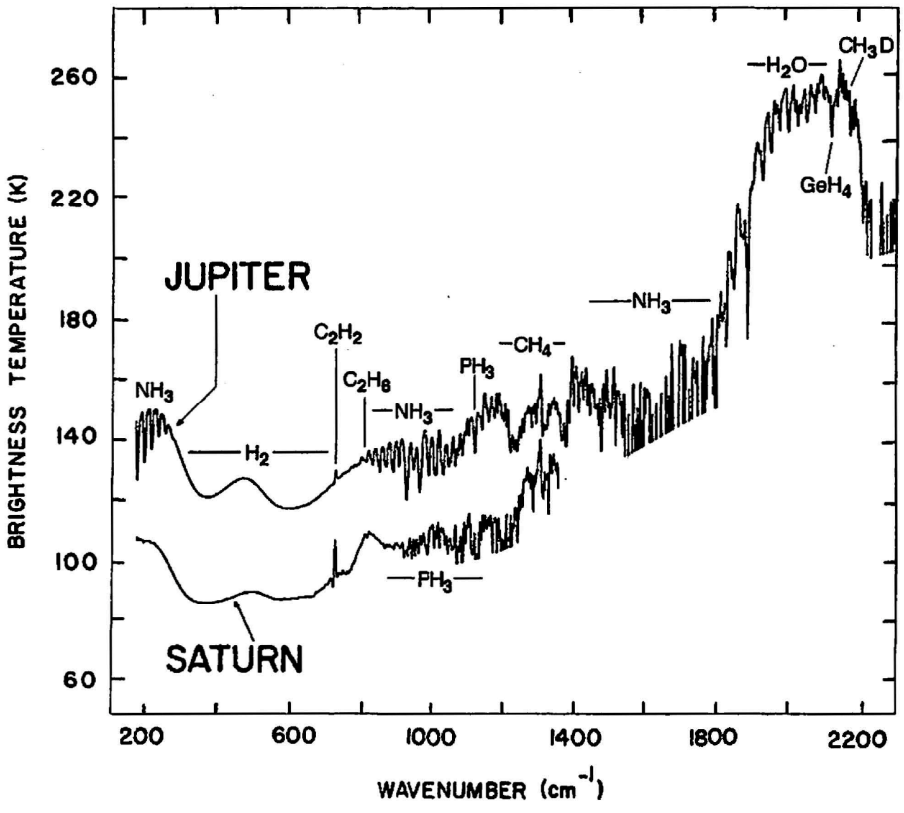
product of the production of methamphetamine drugs [27]. Inhalation can result in death and, even in low quantities, long term exposure may cause anaemia, bronchitis and spontaneous bone fractures [179]. Phosphine is not a molecule that humans should be exposed to, so the ability to accurately detect it should improve its safe use and containment on Earth.

Phosphine in the Solar System

There have been multiple predictions and a few detections of phosphine in astrophysical environments, but its presence and formation scenarios remain poorly understood. Phosphine has so far been detected in Jupiter, Saturn [135, 47, 51, 159, 26] and the carbon star envelope IRC +10216 [1].

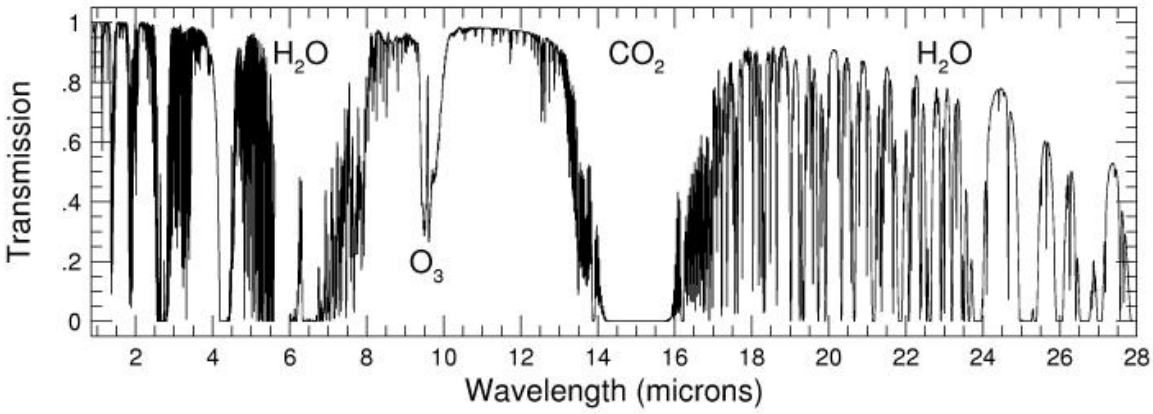
In the early '80s, Voyager provided spectral data on the atmospheres of the gas giants [38, 80], detecting phosphine, amongst other molecules, in both planets. Figure 1.6 compares the spectra of Jupiter and Saturn obtained with the infrared interferometer of Voyager in the wavelength region 100 - 2300 cm^{-1}.

Figure 1.6: Spectra of Jupiter and Saturn recorded with the IR interferometer of Voyager. Figure reproduced from David Catling's notes from the University of Washington course on Planetary Atmospheres ASTR 555.



The 3.2-4 μm spectral region of Jupiter and Saturn is dominated by CH_4 strong bands [122] but the 4-5 μm window, partially due to the lack of strong absorption from CH_4, is a region of low opacity in Jupiter, Saturn and the Earth's atmosphere (see figure 1.7).

Figure 1.7: Infrared absorption of the Earth's atmosphere from 1 to 28 μm showing observation windows, including the 4-5 μm region. Figure taken from "An Introduction to Solar System Astronomy" by Prof. Richard Poggeplots using data calculated by Steven Lord and provided by the Gemini Observatory.



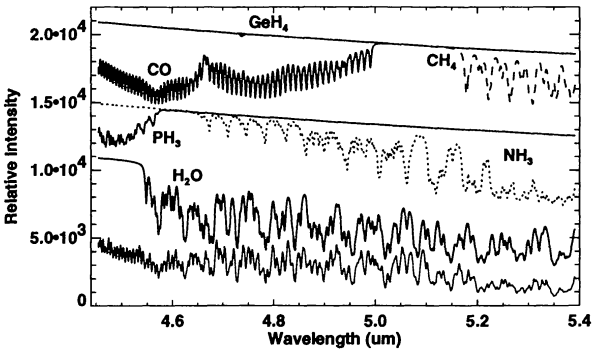
A flux enhancement in this 4-5 μm window is a feature found in many other bodies, from gas planets to brown dwarfs [104]. Consequently CO, GeH_4, AsH_3 and PH_3 all have detectable absorption bands in the 4-5 μm window and PH_3 contributes significantly to its spectrum [123], particularly the $2\nu_2$ overtone bending band. As an example, Figures 1.8a and 1.8b show the predicted relative intensity of several molecules alongside the atmospheric models for Gliese 229b and Jupiter [123].

In Jupiter, the Q branch (see Section 2.3.3) of the $2\nu_2$ phosphine band has been observed in the North Equatorial Belt and the Great Red Spot at1972 cm^{-1} [45]. Far infrared phosphine features predicted by models at 141 μm (71 cm^{-1}) and 103 μm (97 cm^{-1}) have been detected in Jupiter using the long-wavelength spectrometer aboard the infrared space observatory (ISO) [26]. In some regions, e.g. 600-1100 cm^{-1}, phosphine is obscured by ammonia, and it is very difficult to detect any phosphine features [51].

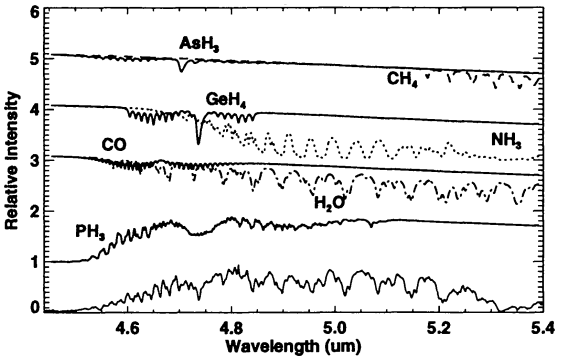
Phosphine plays an important role in Jupiter and Saturn as a marker for vertical and meridian motions and both planets exhibit enhanced phosphine abundances at the equator and depletion in neighbouring belts [51]. In both Jupiter and Saturn, PH_3 has observed

Figure 1.8: Modelled relative intensity curves for CH_4, NH_3, H_2O, CO, GeH_4, AsH_3 and PH_3 in the 4-5 μm window of Gliese 229b and Jupiter. Calculated model spectra represented by the lower solid curves. Figures taken from Noll & Marley [123].

(a) Gliese 229b



(b) Jupiter



abundances which are several orders of magnitude higher than those predicted by thermodynamical equilibrium because chemical equilibrium timescales are long when compared to convective timescales [123]. Deep levels of the atmosphere of Jupiter and, by analogy, brown dwarfs like Gliese 229B, approach an adiabatic temperature profile because of efficient convection [104]. Calculations by Marley *et al.* [104] indicate that the atmosphere of Gliese 229 B then becomes sub-adiabatic* at temperatures below 1700 K, so vertical mixing in these regions would become significant. This is further discussed in chapter 1.3.2.

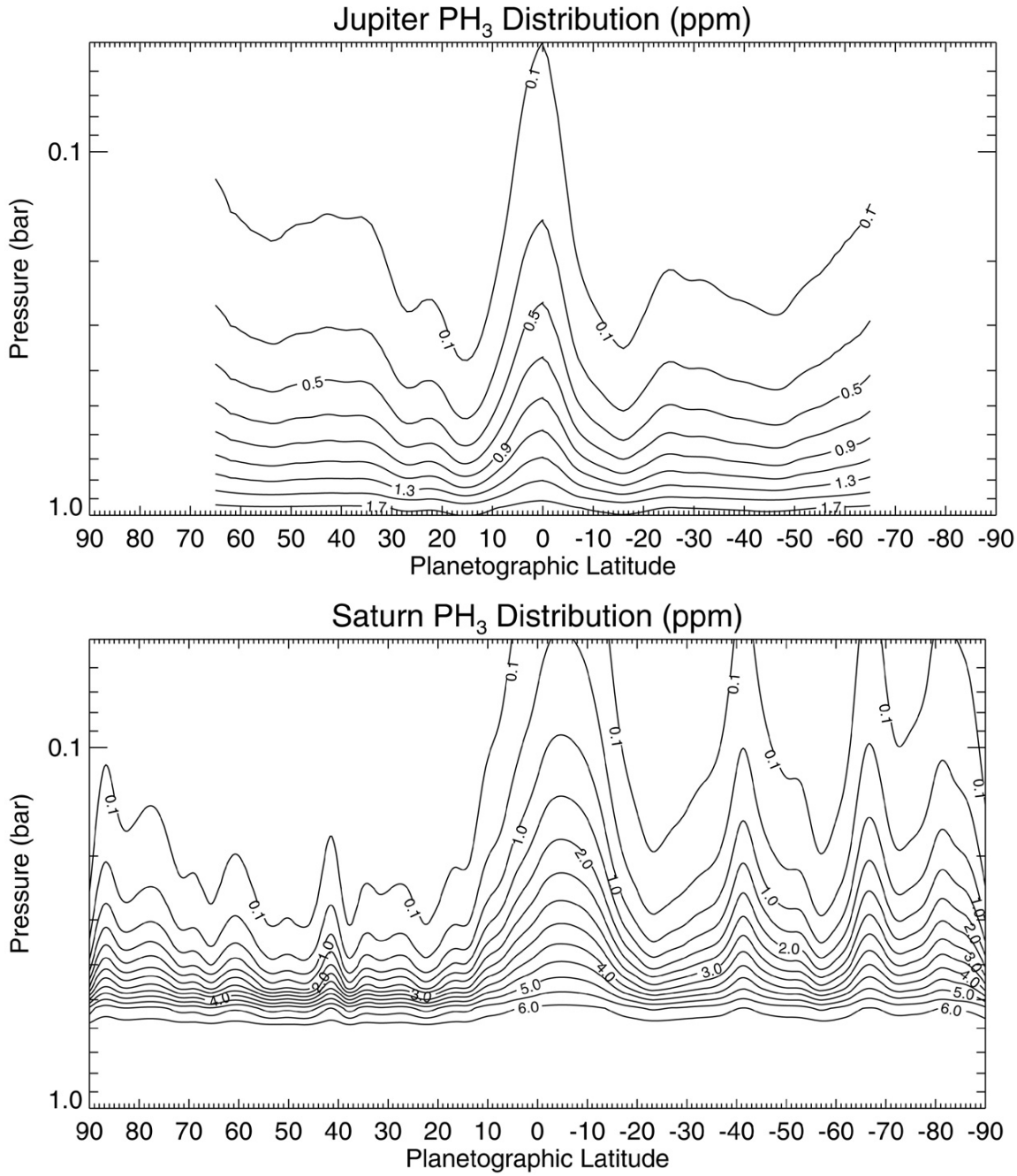
Phosphorus has been found, in the form of phosphine, at approximately tropospheric solar abundances (1.1 ppm) in Jupiter, and is over four times more abundant in Saturn (4.5 ppm) [123, 129, 109, 173]. More recent work by Fletcher *et al.* [51] estimates these values to be much larger, with a phosphorous/hydrogen(P/H) enrichment of 4.8$\pm$0.3 ppm and 15.9$\pm$0.3 ppm above solar composition on Jupiter and Saturn, respectively. This estimation agrees with models of planet formation that expect Saturn to have a considerably larger fraction of ice to gas than Jupiter [83].

Water does not contribute strongly to the spectrum of Jupiter because its upper atmosphere is colder than the condensation temperature of water [110]. In both planets, this abundance of a disequilibrium species like phosphine at observable levels of the atmosphere reflects both a large reservoir in deep regions of the planets and the strength of the convective

*In sub-adiabatic environments, temperature increases with depth more slowly than would be predicted assuming adiabaticity. If hotter regions are also denser, this causes them to sink.

transport of the molecule. Figure 1.9 shows the phosphine abundance profile of Jupiter and Saturn as estimated by Fletcher et al [51] using Cassini/CIRS spectral data.

Figure 1.9: Estimated PH_3 abundance at all altitudes in Jupiter and Saturn based on a pressure knee of 1 bar and deep abundance of 1.9 ppm. Figure taken from Fletcher et al [51].



Analysis of future observations of Jupiter from NASA's TEXES [82], NASA's JUNO [59] and ESA's JUICE [43] will require accurate data on phosphine. JUNO will cover the full latitude and longitude of Jupiter and its analysis of the polar atmosphere of Jupiter should confirm a depletion of phosphine in the polar vortices analogous to Saturn's [48, 51]. Past observations of Saturn from Cassini/VIMS and Cassini/CIRS will also gain from further understanding of the PH_3 spectrum. Suspected inaccuracies found in regions of the existing phosphine data may be behind misinterpretation of previous astronomical observations [103].

Titan has also been predicted to contain phosphine, but spectra recorded with Cassini CIRS failed to detect any, putting an upper limit of 1 ppb on the phosphine abundance of its atmosphere [120]. Similar unsuccessful detections of phosphine in Neptune and Uranus suggest that the abundance of phosphorous on these planets is sub solar [109], with an upper limit of 10% solar P/H abundance.

Phosphine Outside the Solar System

Phosphine has recently been confirmed by the HIFI instrument on board Herschel in the carbon star envelope IRC +10216, with an abundance of 10^{-8} relative to H_2 [1]. The observed PH_3 lines in this circumstellar environment had a width consistent with the terminal velocity of the star's envelope, suggesting phosphine is emitted from regions where the gas had reached this velocity. It was also concluded that infrared pumping to excited vibrational states has an impact on the excitation of phosphine in the envelope; this increases the excitation temperature of rotational transitions and hence their line intensities. Together with HCP, PH_3 is one of the major phosphorus carriers in the inner circumstellar regions.

By analogy to Saturn and Jupiter, it is expected that PH_3 could be used as a marker for vertical convection and a tracer for tropospheric dynamics in the upper atmospheres of other astronomical bodies, namely Hot Jupiters and Brown Dwarfs. For example, PH_3 should be detectable (for resolutions higher than 1.2 cm^{-1}) in the 4.45-5.40 μm spectral window of Gliese 229B, where it is predicted to be the dominant phosphorous molecule [123, 49].

There have been multiple predictions of phosphine in many other astrophysical environments, but its presence and formation scenarios remain poorly understood and no phosphine has been found outside the Solar System other than the aforementioned discovery of circumstellar phosphine in IRC +10216. This is due to a variety of factors, including the sparsity of molecular data which this work is rectifying.

1.3.1 Phosphine in Chemical Equilibrium

When considering local thermal equilibrium (LTE) and solar metallicity, phosphorous can be expected to be found in the form of P_4O_6, PH_3, P_2, PH_2, PS, PO, or in its elemental form, P [173]. As can be seen from Figure 1.10, P_4O_6 dominates at low temperatures, and as temperature rises above 1000 K, PH_3 and P_2 become the dominant phosphorous species.

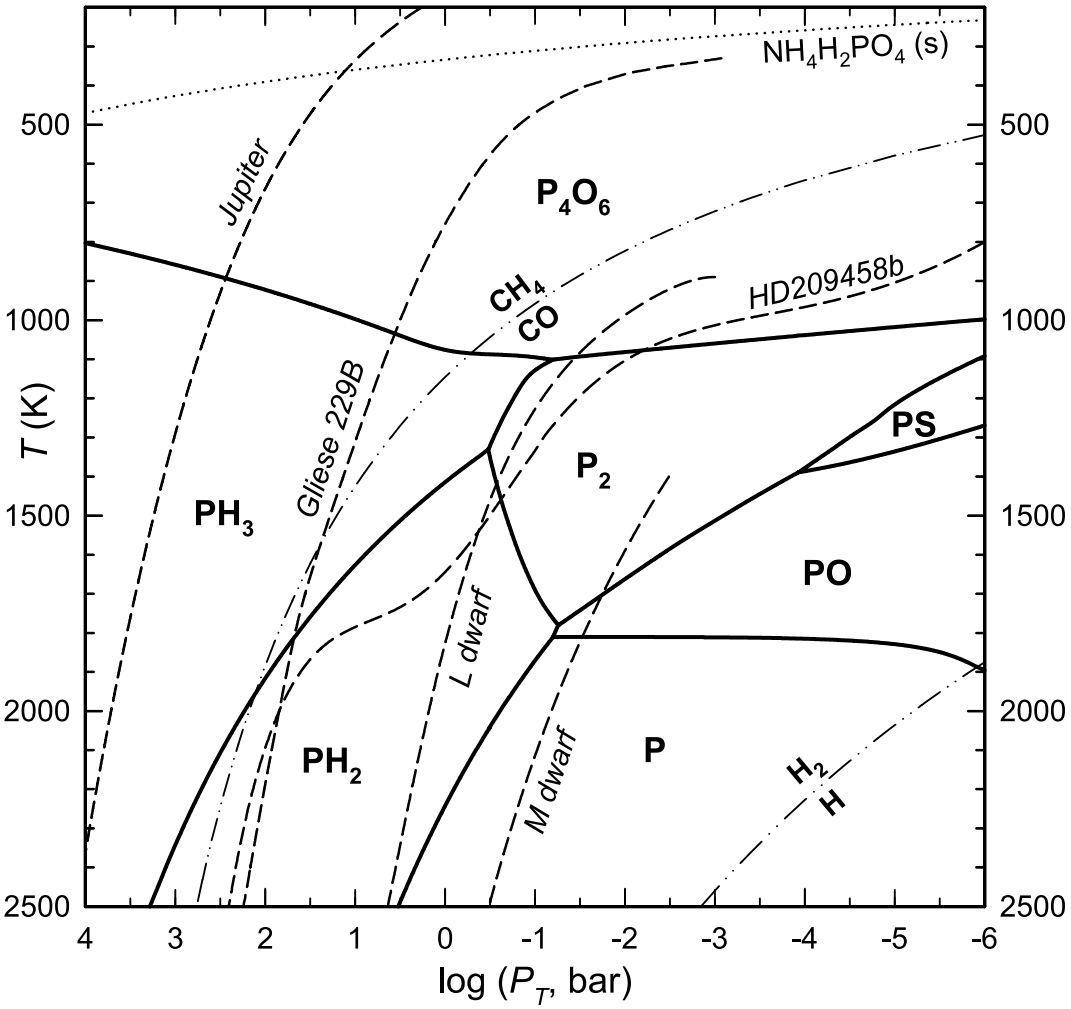


Figure 1.10: Phosphorus chemistry as a function of temperature and pressure. The solid lines represent equal molecular abundances. Dotted line shows the condensation curve for $NH_4H_2PO_4$ and the dash-dotted lines equal abundance boundaries. Also present are the atmospheric profiles for a few types of substellar objects. Figure reproduced from Visscher et al. [173].

The triple point of PH_3-P_4O_6- P_2 is predicted to be at a temperature of 1101 K and pressure of $10^{-1.2}$ bar, and the relative importance of the P_4O_6-P_2 and the P_4O_6-PH_3 reaction depends on where an object is located on the P-T profile. At higher total pressure, phosphine is the dominant phosphorous carrier but, at temperatures over 1000 K, phosphine will not

form if the pressure decreases sufficiently, and phosphorous will be found as P_2, PH_2, PS or PO. Figure 1.10 also shows atmospheric profiles for a few substellar objects, in dashed lines. It can be seen that, in L dwarfs, PH_2 and P_2 are expected to be present in similar amounts when in LTE.

The abundance of phosphine in chemical equilibrium phosphine varies as a function of temperature and pressure, which is shown in Figure 1.11, assuming solar metallicity. It can be seen that for high temperatures ($T \geq 1000$ K), PH_3 abundances generally increase with pressure but decrease with temperature, when other P-bearing gases become dominant.

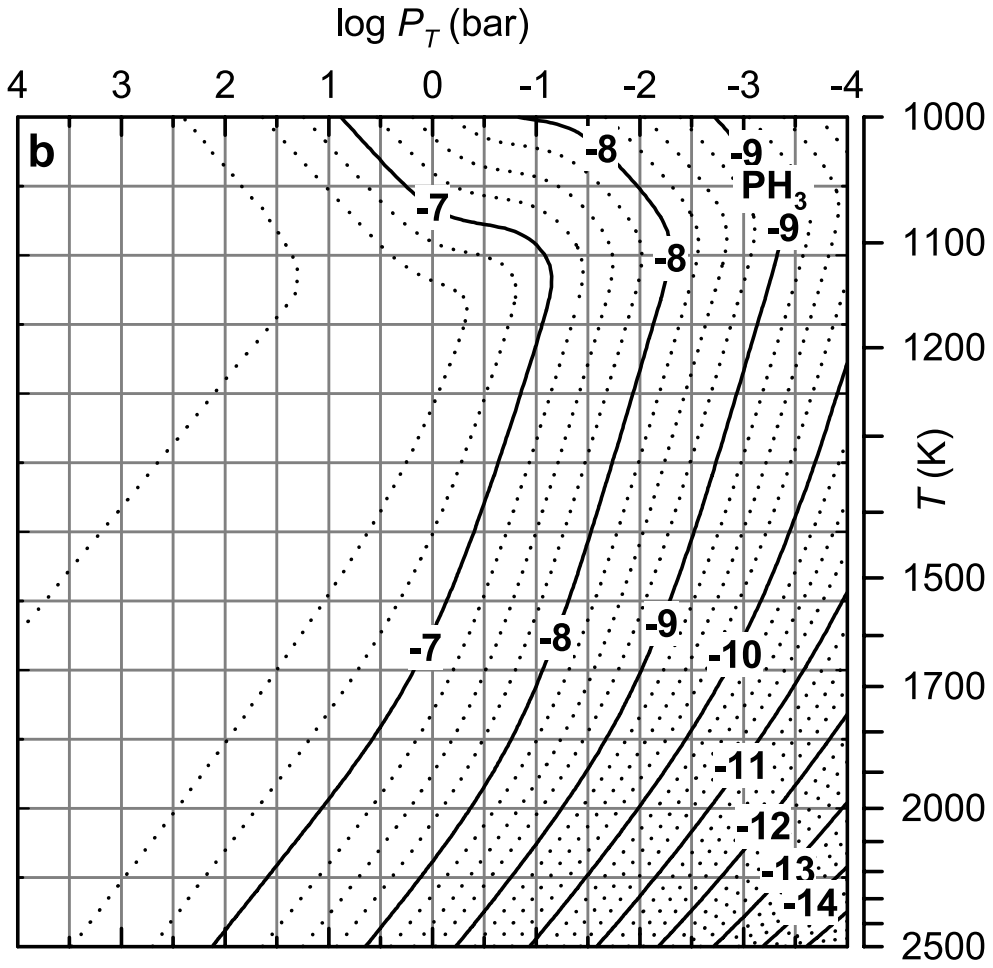


Figure 1.11: Mole fraction contours (on a logarithmic scale of 10^4) for phosphine as a function of pressure and temperature in a solar metallicity gas. Figure reproduced from Visscher et al. [173].

When compared to the relative abundances of other phosphorous chemistry in equilibrium, it appears that phosphine contains the entirety of the atmosphere's phosphorous in the deep layers of T dwarfs and giant planets.

Figure 1.12 (also taken from Visscher et al [173]) illustrates the phosphorous chemistry along a P-T profile for models of the atmospheres of Jupiter, Gliese 229B, HD 209458b and a general L dwarf.

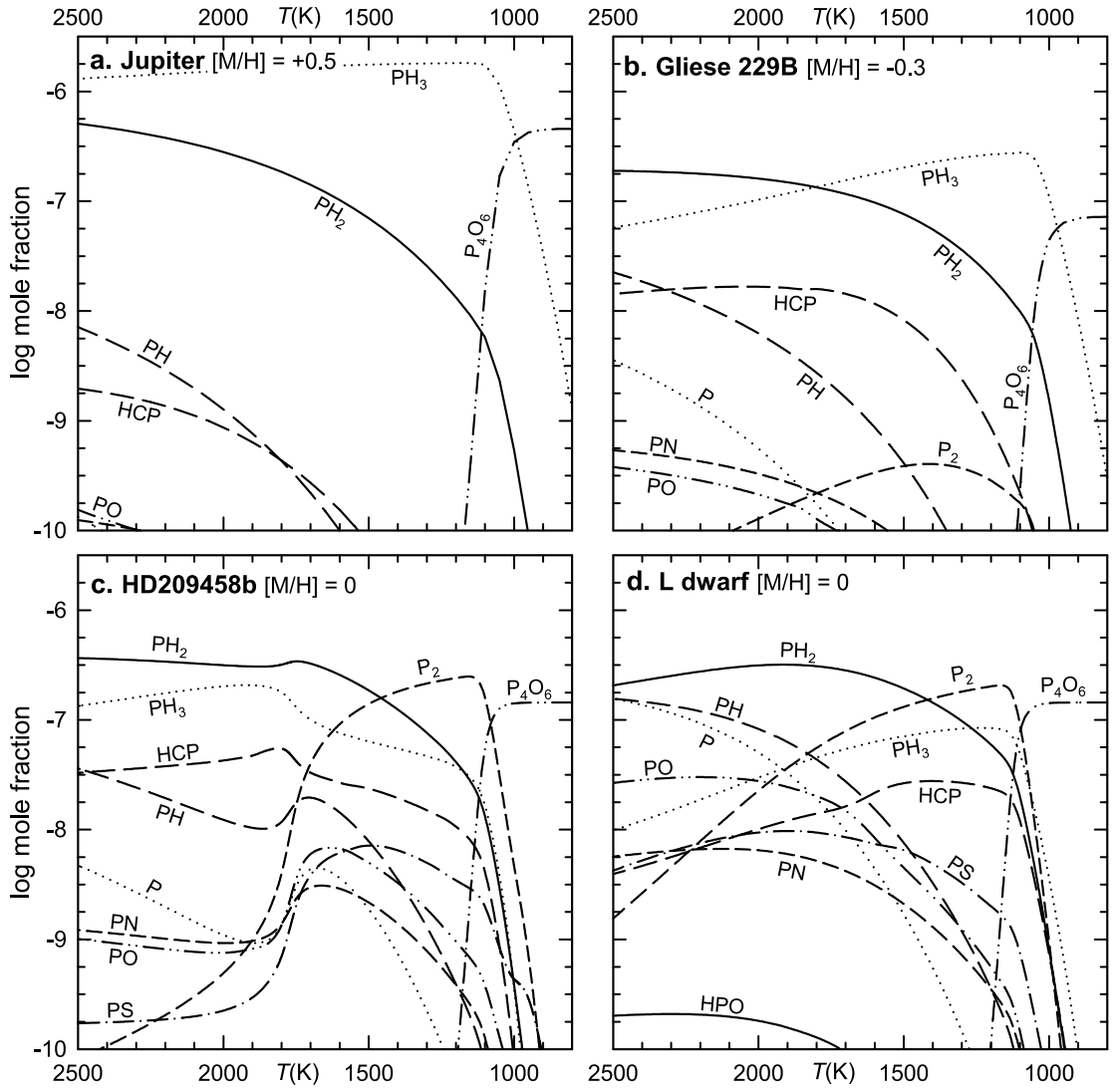


Figure 1.12: Phosphorus equilibrium chemistry along pressure and temperature for (a) Jupiter (T $_{eff}$ =124 K), (b) Gliese 229B (T $_{eff}$ = 960 K), (c) HD 209458b (T $_{eff}$ = 1350 K), and (d) an L dwarf (T $_{eff}$ = 1800 K). Figure reproduced from Visscher et al. [173].

At thermal equilibrium, phosphine is expected to be the dominant phosphorous carrying species at the observable atmosphere of hot T dwarfs and cool L dwarfs, with effective temperatures between 1000 and 1400 K [173], while P_4O_6 would be expected to dominate the upper atmospheres of giant planets and other T dwarfs. However, almost no P_4O_6 is found in these environments probably because chemical equilibrium is not established.

1.3.2 Disequilibrium Chemistry

In very hot environments, such as the dayside of hot exoplanets and brown dwarfs with T $\geq$ 2000 K, the observed chemical composition can be expected to follow thermochemical equilibrium conditions. However, at temperatures below this, and at the uppermost atmosphere of most objects, the kinetic dynamics of an environment play a very important role on its composition, and hence its radiative properties and thermal structure [111]. One of these processes is transport induced quenching. Quenching occurs at temperatures where the rate of chemical reactions approach the timescales for chemical transport. When this occurs, the local atmospheric composition deviates from chemical equilibrium due to convection or other large scale diffusion. In the case of PH_3 in Jupiter and Saturn, the rate of photochemical destruction in the planets' upper troposphere approaches that of the rate of vertical mixing from within deeper layers of the atmosphere, so the phosphine inventory at the cloud-top is constantly being resupplied by violent mixing from the deep layers of the atmosphere [51].

In the PH_3 to P_4O_6 oxidation, the quenching temperature occurs inside the PH_3 field in the deep layers of T dwarfs and giant planets. This causes phosphine to mix vertically towards the observable atmosphere without P_4O_6 having time to form. Even at low temperatures, where P_4O_6 is often the dominant phosphorous-containing gas under equilibrium conditions, rapid vertical mixing and phosphine quenching from regions where PH_3 is dominant predict high disequilibrium abundances in the upper layers of the atmosphere [173]. For example, the phosphine abundances mentioned above for Jupiter and Saturn are approximately thirty orders of magnitude larger than that expected from thermodynamic equilibrium predictions; this is evidence for rapid vertical quenching, and it is expected to be an effect replicated in other giant planets. If the quench temperature of phosphine is at an altitude where it is the dominant phosphorous species, it is likely that phosphine will be vertically mixed up through the atmosphere, to an altitude where UV photochemical destruction begins to occur [51]. In many disequilibrium scenarios, the entirety of the phosphorous inventory is expected to be represented as phosphine, as the conversion of PH_3 to P_4O_6 is kinetically inhibited. Phosphine is then mixed sufficiently upwards for its detection to be possible in the observable atmospheres of giant planets and several types of dwarfs [173, 49].

1.3.3 Predictions for Phosphine Detection

Given its importance in both LTE and non LTE environments, it is predicted that phosphine contributes strongly to the spectral profile of many substellar objects. Below is a summary of environments where phosphine is expected to be detected.

In Y dwarfs, with $T_{\text{eff}} \approx 500$ K, phosphine should exhibit a strong feature in the mid-infrared at 4.3 μm [110], where it is predicted to be the dominant source of opacity. The spectra of most T dwarfs, with $T_{\text{eff}} \leq 1300$ K, will be dominated by other molecules (e.g. H_2, NH_3, CH_4, CO, H_2O and CO_2) but phosphine can still have a significant contribution to the shape of the spectrum. The NIRSpec near-infrared spectrograph on JWST will be able to collect spectra in the 2.9-5 μm region with sufficient sensitivity to detect the presence or absence of phosphine, particularly for cooler Y dwarfs with $T_{\text{eff}} \leq 450$ K [110]. In cool T dwarfs, like Gliese 229B, phosphine is predicted to be the dominant phosphorus molecule where it is expected to carry all the phosphorus in the atmosphere to the top layers, with approximately 0.6 ppm. Phosphine should be detectable (for resolutions higher than 1.2 cm^{-1}) in the 4.45-5.40 μm spectral window of Gliese 229B, particularly the 4.3 μm feature [123, 49]. Models estimate that, for regions where $T \geq$1155 K, the major phosphorus carrying gas in Gliese 229B is phosphine, with rapid vertical mixing quenching its destruction [49]. As the temperature rises, PH_3 is converted to P_4O_6, but it should still be present at observable heights. These calculations consider electroneutrality, mass balance, and chemical equilibrium constraints.

In the deep layers of the atmospheres of T dwarfs and hot jupiters, PH_2 is the dominant phosphorous species and, as the temperature decreases, is replaced by PH_3 and/or P_2, depending on the position of the PH_3-PH_2-P_2 triple point (see figure 1.10) [173]. In hotter environments, like the exoplanet HD209458b and L dwarfs, where phosphorus equilibrium chemistry is approached, the phosphine abundances decrease significantly to about 50 ppb [173]. It should still be possible to detect phosphine in these atmospheres, but higher resolutions may still be required to distinguish its features from that of neighbouring CO bands at low wavelengths [123].

The detectability of phosphine is also strongly influenced by gravity, with lower gravity models showing increased PH_3 absorption in the 3.5-4.7 μm spectral region [110]. Since this is a spectral window between strong water and methane absorptions, phosphine could have a large contribution to the shape of the spectra for bodies with lower mass.

There are still disequilibrium effects that are not completely understood. For example,

observations of Saturn's upper troposphere found that phosphine abundances decreased with depth [103], contradicting the proposal that phosphine is dredged up through vertical mixing from deeper layers of the atmosphere. This depletion is not understood, as there is no known formation mechanism for phosphine at the upper levels of Saturn's atmosphere that could lead to the abundances observed. Suggested explanations for this inconsistency are a poor understanding of the aerosol opacity in the VIMS observations and intensity uncertainties in the molecular data.

To continue modelling and investigating the environments described in this chapter, and to ensure the correct detection of phosphine in future observations, it is crucial to have access to a comprehensive, high resolution, line list for phosphine. The ultimate aim for the work described in this thesis is to create such a line list. The corresponding spectrum can then be extrapolated over a wide range of temperatures, pressures and path lengths. Only then can it be possible to identify and interpret phosphine features accurately in astrophysical spectra.

The thesis builds on previously released ExoMol line lists for polyatomic hydrides such as water [14], ammonia [183], formaldehyde [4] and methane [190, 191].

The next chapter sets up the key theoretical background for the understanding of line lists.

Chapter 2

Theoretical Background - From Astronomical Light to a Line List

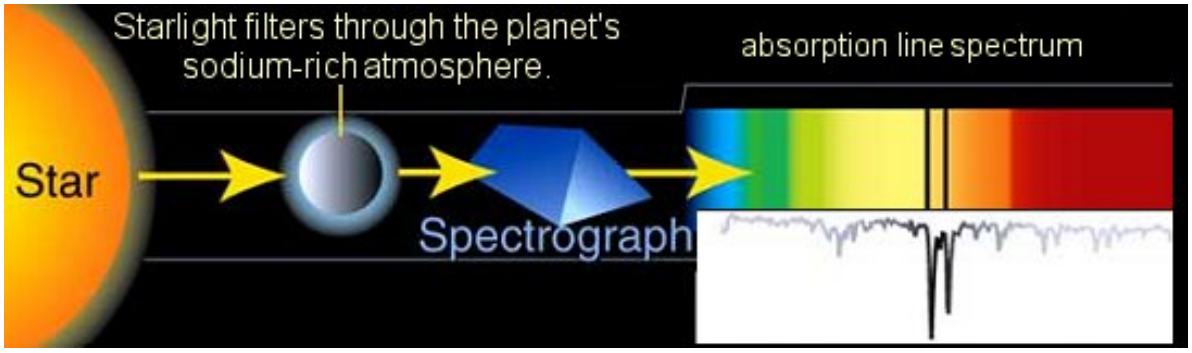
Extracting information from astronomical spectra is a complex operation. As previously discussed in section 1, accurate observations are difficult to make, and it is currently impossible to retrieve astronomical spectral data with enough resolution and SNR to characterise the atmospheres of Earth like planets. Once the spectrum of an exoplanetary atmosphere is obtained, interpreting its composition is far from trivial, as many factors(mixing ratios, cloud coverage, etc) affect its appearance (see, for example, [104, 91, 7, 168, 145, 39, 138]). A crucial part of this process is the existence of accurate, complete, molecular line lists. An overview of the journey between light spectra to molecular composition is described in this chapter, which is the driving force for the computation of molecular line lists. Additionally, this chapter will describe the general methodology behind the creation of the line lists themselves.

2.1 Light Spectra

Light spectra can be characterised as a continuous intensity distribution with multiple drops and spikes, corresponding to absorption and emission lines, respectively. These lines can correspond to the fingerprint of atoms, ions or molecules. Molecules are only found in regions of the universe which are cool enough to sustain them, such as planetary atmospheres, brown dwarfs, comets, interstellar clouds and starspots. Polyatomics in particular can only be found in any significant quantity in environments below 4000 K [160].

By examining the spectra of a variety of astrophysical bodies, many detections of molecules have already been made in brown dwarfs, M-dwarfs, supernova remnants, comets and many more astronomical bodies (see for example, [126, 195, 19, 66]). The first atom to be detected in an exoplanet atmosphere was sodium [153], which has a very distinct spectral signature. Figure 2.1 illustrates this process.

Figure 2.1: Illustration of the detection of sodium in the atmosphere of the exoplanet HD 209458b. Figure credited to *A. Feild, STScI and NASA*.



Despite many detections of molecules within and outside our Solar System, molecular detections in exoplanetary atmospheres are still mostly restricted to H_2O, CO, CO_2 and CH_4 (see, for example, [167], [157], [156] and [73]). The wealth of molecules present in the Earth's atmosphere and throughout the interstellar and circumstellar media suggests that hundreds of molecules remain undetected in the atmospheres of exoplanets. The comparative scarcity of molecular detections in exoplanet atmospheres is due partially to the relative youth of the field, but also to inaccurate modelling, the difficulty in obtaining good atmospheric data, and ultimately the paucity of high quality molecular data. Figure 2.2 provides an overview of the molecules believed to be detectable in astrophysical atmospheric characterisation, roughly distributed by environment, as well as their availability.

When light goes through an environment where molecules form, these absorb part of the light spectrum. Since every atom and molecule absorbs at only very specific wavelengths in the spectrum, each component can be identified by its spectral fingerprint. This fingerprint arises from the allowed states of a molecule which have very specific, discrete values. As a consequence, light that travels through an environment will also be absorbed only at the very specific, discrete frequencies that correspond to that environment's atomic and molecular composition. If the exact fingerprint of each species is known, then its presence can

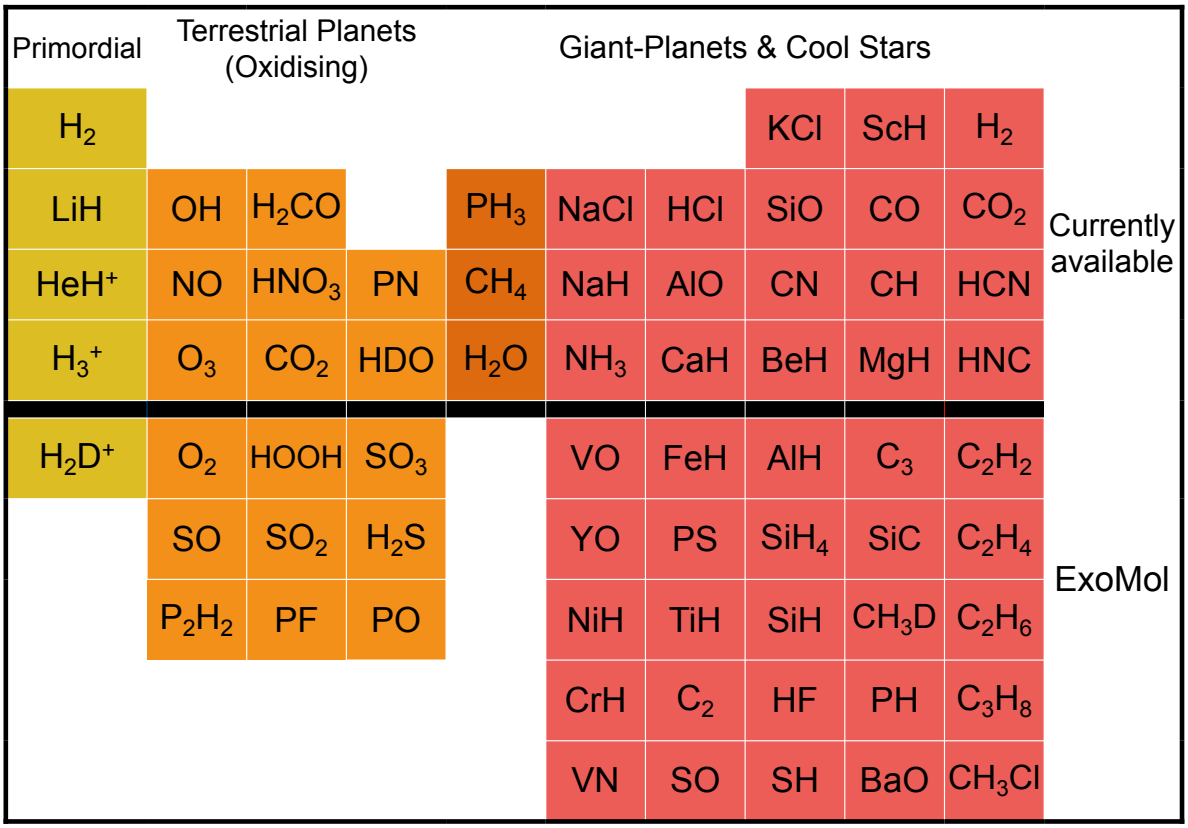


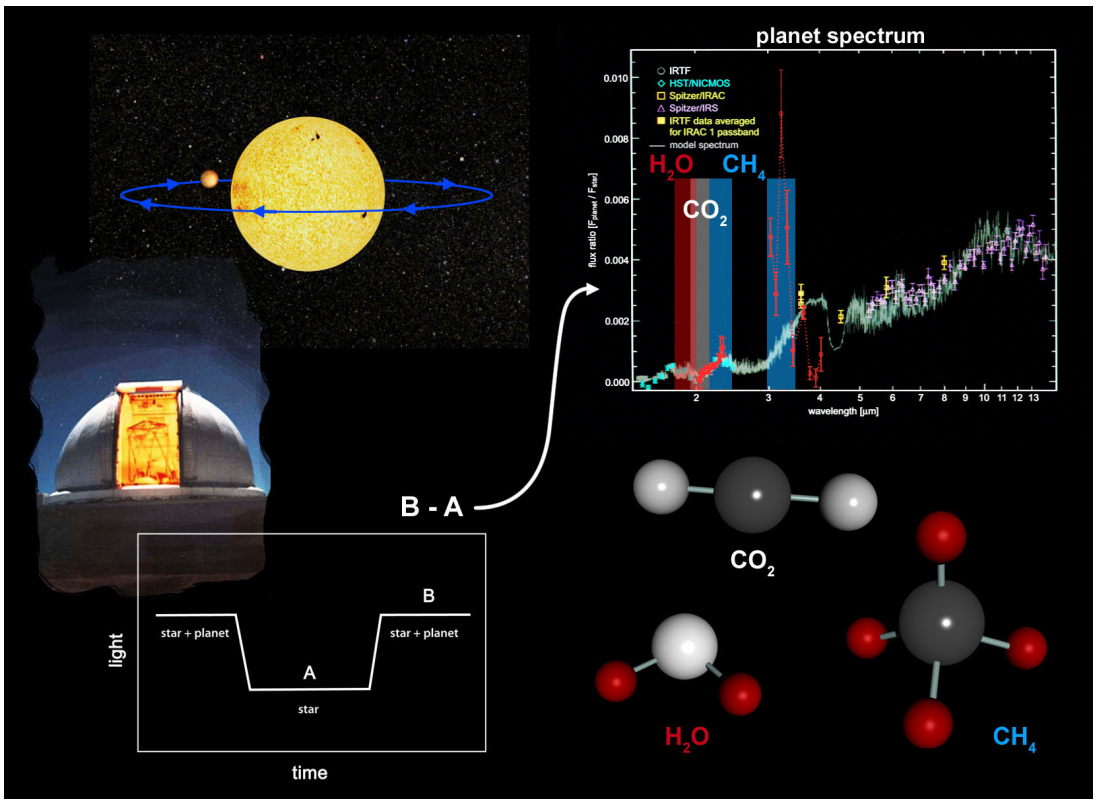
Figure 2.2: Overview of predicted and detected molecules in different astrophysical environment. Most of the central molecules could be present in multiple environments. The molecules in the lower half of the figure correspond to molecules that ExoMol project is currently, or will soon be, studying [165].

be confirmed from that light's spectra. Likewise, it should be possible to theoretically predict what a spectrum of a body should look like if we know what molecules it contains, its thermodynamical and physical conditions, and how the molecules themselves behave in such an environment. Ultimately, knowledge of the spectral fingerprint of each molecule that contributes to the spectrum of a body is crucial, whether one is performing an accurate analysis of an observed atmosphere or an accurate prediction of what that body's atmospheric spectrum should look like.

As previously mentioned in chapter 1, to obtain a spectrum from an exoplanet atmosphere, the spectrum of the host star will often have to be isolated and subtracted from observations, so that the absorption and emission features for which the exoplanet is exclusively responsible can be identified.

It is extremely difficult to obtain astrophysical spectra with the required accuracy to analyse it unambiguously; if it is done, then an atmospheric or radiative transfer model can incorporate molecular and atomic spectroscopic data in an attempt to replicate the observed spectra. However, these spectra are often extremely dense and rich in structure and, if containing multiple, complex molecules, they can be extremely hard to interpret. Nonetheless, with a sophisticated understanding of molecular behaviour, together with fundamental data for all the species that contribute significantly to an atmospheres' opacity, it is possible to combine observational data and accurate retrieval models to understand the composition and structure of an exoplanet atmosphere. Figure 2.3 illustrates how transit spectroscopy can be used to confirm the presence of multiple molecules in an atmosphere, using the example of the hot jupiter HD 189733b.

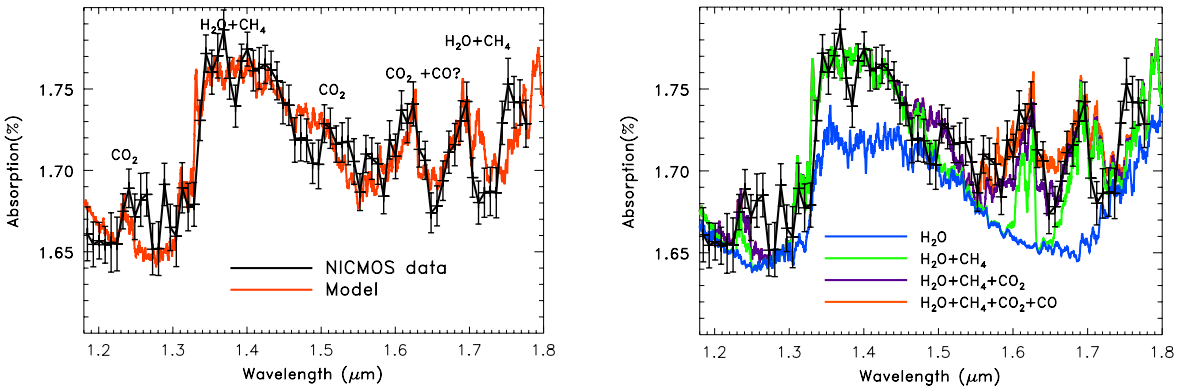
Figure 2.3: Illustration of the detection of water, carbon dioxide and methane in the atmosphere of the exoplanet HD 189733b. Figure credited to *NASA/JPL-Caltech*.



This hot jupiter is of particular interest because both the first detection of water [167] and the first detection of methane [157] in an exoplanet atmosphere were made for HD 189733b. The identification of water was done using a theoretical, high resolution, complete water line list, which was the precursory line list to the ExoMol project [14].

If the fundamental molecular data exists then computational models can verify which combination of molecular abundances and thermodynamic parameters best fit the observed atmospheric spectra. Figure 2.4 demonstrates the effect of including additional molecules when modelling an atmosphere [166]. The image on the left shows the best fit to the data, although the absence of an accurate, complete CH_4 line list at the time of modelling made it difficult to confirm the presence of CO, as the methane absorption at this wavelength range was not fully understood when the associated paper was released [190].

Figure 2.4: Primary transit spectrum of the Hot Jupiter XO-1b, obtained with the Hubble-NICMOS instrument. The observational data points are in black, and the atmospheric models containing H_2O, CH_4, CO, and CO_2 are shown as coloured lines. Figure credited to *Tinetti et al* [166].



Observational data from exoplanet atmospheres will be a combination of light spectra and noise from instrument systematics, stellar activity and other sources. Separating the original light curve from the noise in the data is a very challenging process, discussed at length in a vast number of research papers (see, for example, [175]). Nonetheless, correcting for noise in the data is one of the most crucial stages for the extraction of a clear light curve from observations. Once an accurate light curve is obtained, the process of extracting molecular information from it can begin.

A light curve can be described by the Beer-Lambert law as a frequency dependent intensity, or $I(\nu)$, which can be defined as:

$$I(\nu) = I_0(\nu)e^{-c\sigma(\nu)l} \tag{2.1}$$

where the light received, $I(\nu)$, is dependent on the original light curve, $I_0(\nu)$, the path length

of the medium, l, as well as the number concentration, c, and cross-section σ of the species in the environments that the original light travelled through.

In transit spectroscopy, $I(\nu)$ and $I_0(\nu)$ can be considered as the starlight at primary eclipse and at secondary eclipse, respectively. The log of their ratio, known as the flux ratio or transmittance, then gives:

$$\ln\left(\frac{I_0(\nu)}{I(\nu)}\right) = \sum_k c_k \sigma_k(\nu) l \tag{2.2}$$

where k accounts for each species in the medium.

The path length of the medium, l, can be retrieved from geometric models and the flux ratio can be obtained from observations. However, $c_k\sigma_k(\nu)$, known as opacities, or absorption coefficients, are harder to obtain. The number concentration, c, is sometimes also called number density, and it is equivalent to Nx_k, where x_k is the mixing ratio of a particular molecule, and the N is the total number density of molecules. The number concentration, c or Nx_k, consists of parameters that are model dependent and usually are inferred from existing predictions or knowledge of the astrophysical environment in question. c is usually quoted at standard temperature and pressure.

The only parameter left to obtain from Eq. 2.2 is the frequency-dependent absorption cross-sections, $\sigma_k(\nu)$, or, if wavelength dependent, $\sigma_k(\lambda)$. It is worth noting that the standard units for the Eq. 2.2 are cm^2/molecule for $\sigma_k(\lambda)$ and molecule/cm^3 for c_k which, together with the path length, l, makes for a unitless value, as would be expected from a flux ratio.

The absorption cross-sections, $\sigma_k(\lambda)$ or $\sigma_k(\nu)$, must be provided for each individual species, and they are not only frequency dependent but also vary with pressure and temperature. For each species, k, one can consider σ_k to be the sum over the contributions from individual cross-sections, σ_{ij}, of all the relevant molecular transitions of that species within the bin under consideration [71]; the number of transitions included in a cross-section of a specific bin-size, $\Delta\widetilde{\nu}$, will vary with resolution.

Each cross-section, σ_{ij}, which contributes to a range of frequencies $\Delta\widetilde{\nu}$, within a grid, $\widetilde{\nu}$, can be defined as:

$$\sigma_{ij} = \frac{I_{ij}}{\Delta\widetilde{\nu}} \int_{\widetilde{\nu}-\Delta\widetilde{\nu}_i/2}^{\widetilde{\nu}+\Delta\widetilde{\nu}_i/2} f(\widetilde{\nu}; \widetilde{\nu}_{0;j}, \alpha_j) d\widetilde{\nu} \tag{2.3}$$

where I_{ij} is the line intensity for each transition, from i to j, and $f(\widetilde{\nu}; \widetilde{\nu}_{0;j}, \alpha_j)$ is the line function of the cross-section. The cross-section is then dependent on the value of the integral

over the line shape function within that frequency space, where $\widetilde{\nu}_{0;j}$ is the line centre position and α_j is the Doppler half-width at half-maximum (HWHM).

The line function itself depends on a few factors. Each spectral line will have a natural broadening of finite width, caused by the natural lifetime of the upper state. A typical lifetime of an isolated molecule will be $\approx 10^{-8}$ s, corresponding to a 5×10^{-4} cm^{-1} broadening to the line width [85]. Increased pressure and temperature will further reduce lifetime and consequently increase the line broadening. Pressure induced by collisions with other molecules will increase the line width, and will depend on the species present and their number density; this is known as pressure broadening. The temperature, and corresponding relative speed of the molecules, will also affect the line width, with changes in the molecule's thermal velocity causing the incoming radiation to change frequency with respect to the molecule's reference frame; this is called the Doppler effect. If only the Doppler broadening is considered when establishing the line function, a Gaussian line shape is sufficient [71]. If it is necessary to account for the effect of pressure broadening from neighbouring species, a Lorentzian profile is required which, convoluted with a Gaussian function, is called a Voigt profile. This is necessary because the Doppler effect and collisions between molecules cannot be considered as completely independent from each other. Several distributions of Gaussian and Lorentzian parameters in Voigt line shapes can be seen in Figure 2.5.

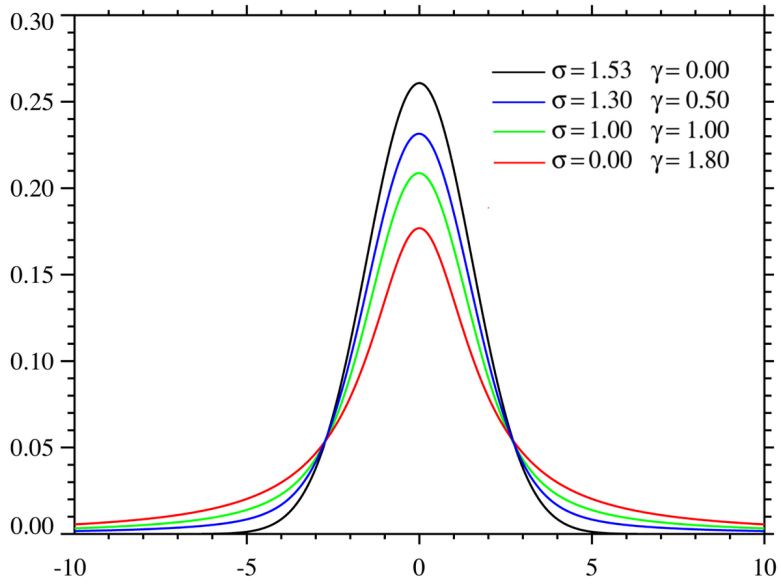


Figure 2.5: Voigt profiles for four different Gaussian (γ) and Lorentzian (σ) contributions. The black and red profiles are the limiting cases of the Gaussian ($\gamma = 0$) and the Lorentzian ($\sigma = 0$) profiles respectively. Figure credited to Wikimedia Commons.

Once a frequency grid is chosen, and the line shape function computed, the only component of the cross-sections left to obtain is the temperature dependent individual line intensity, I_{ij} or $I(f \leftarrow i)$, which corresponds to the intensity of the transition between initial and final states, i and j:

$$I(f \leftarrow i) = \frac{A_{if}}{8\pi c} g_j \frac{e^{-E_i/k_b T}}{Q(T)\ \tilde{\nu}_{if}^2} \left[1 - \exp\left(\frac{-hc\tilde{\nu}_{if}}{k_b T}\right)\right] \tag{2.4}$$

where $I(f \leftarrow i)$ has units of cm^{-1}/(molecule·cm^{-2}), h is the plank constant, k_b is the Boltzmann constant and $Q(T)$ is the temperature dependent partition function, which is discussed at length in chapter 4. Once these constants are obtained, all that remains is the degeneracy of the upper state (discussed further in the next section), g_j, the lower state energy, E_i and the Einstein A coefficient for the transition, A_{if}:

$$A_{if} = \frac{8\pi^4 \tilde{\nu}_{if}^3}{3h} (2J_i + 1) S_{if} \tag{2.5}$$

where J_i is the rotational quantum number for the initial energy state, E_i, and S_{if} is the temperature independent transition line strength.

Together, the line strengths, S_{if}, transition frequencies, $\tilde{\nu}_{if}$, and the energy states between which the transitions occur, effectively describe a transition line list. The creation of such a line list, and the calculation of all its relevant components, is the main purpose of the remainder of this chapter.

2.2 Obtaining a Line List

If one is able to calculate an accurate temperature and pressure dependent line function then, for a specific resolution, the correct assignment of an opacity to an atomic or molecular signature requires temperature dependent spectra, or a transition line list. These can be summarised as a description of possible lower and upper energy states of a molecule, and an understanding of the likelihood of transitions between them.

A hypothetical, perfect atmospheric model, with an exact line profile description, would still fail to accurately replicate observations if the transition line data, and consequent cross-sections, are inaccurate or incomplete. In fact, problems with the quality and/or quantity of molecular line lists is one of the most important limiting factors in our ability to accurately model high resolution spectra for hot substellar objects.

Line lists can be obtained by analysing the experimental spectrum of a molecule at a specific temperature, and describing each resolved transition in the observed range. Using the information gathered from the spectrum it is then possible to predict how each transition will change with temperature, provided the lower energy state has been correctly assigned and the temperature dependent partition function is available. Alternatively, the molecular transitions and corresponding energy states can be calculated from first principles, for any given temperature. In reality, a combination of theoretical and experimental methods is usually required to obtain an accurate, comprehensive line list.

2.2.1 Experimental Concerns

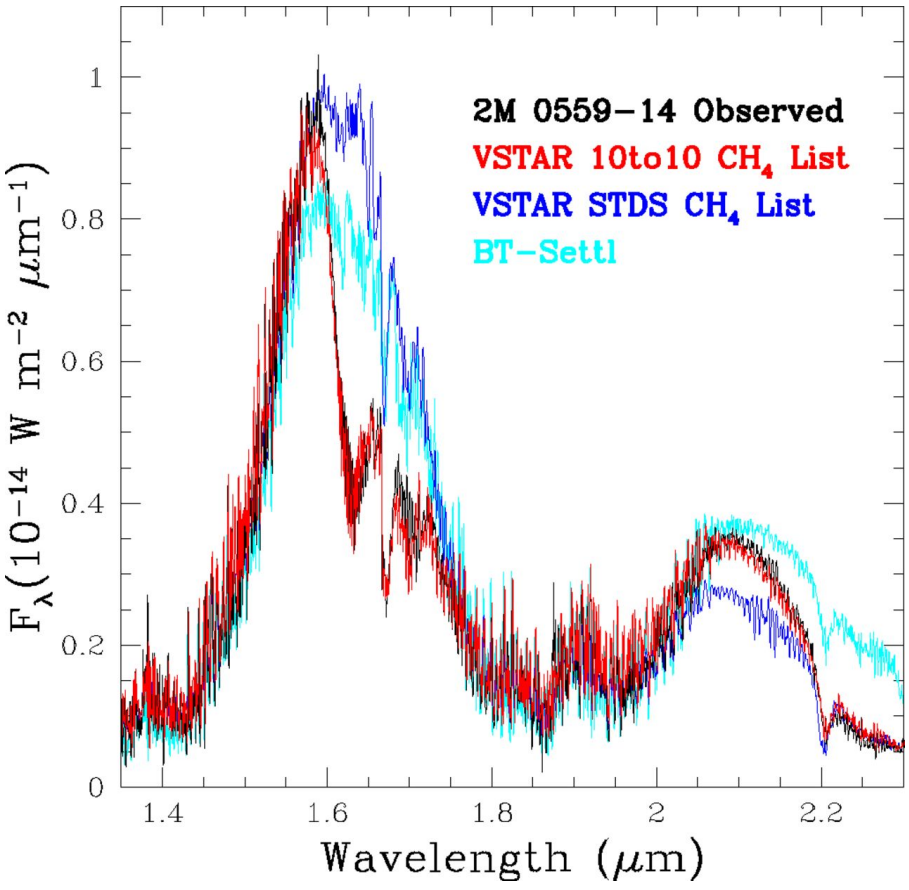
Experimental measurements of line lists were for many decades the only source of spectral knowledge. Measurements of line position can achieve incredible precision and accuracy levels (up to 10^{-8} cm^{-1}), currently beyond the reach of theoretical alternatives [161]. Line intensities are more difficult to measure accurately; the absolute accuracy of intensity measurements will depend on many experimental sources of error, e.g. sample purity, pressure gauges. For example, contaminated gas samples were a likely cause for the systematic overestimation of PH_3 intensities in the Kitt Peak spectra [103], by approximately 10%. Due to their increasing reliability, theoretical predictions of intensity are now beginning to be used to improve or completely replace those in laboratory databases [162, 165, 94, 171]. Additionally, a lot of spectroscopic research that requires laboratory measurements is not particularly concerned by intensity measurements, so some excellent experimental data on line positions is available without any intensity information. In some experiments, absolute intensities are not measured at all, such as some microwave spectra or spectra at high temperatures where the environment is not in thermal equilibrium [162]. For astronomical observations or atmospheric modelling these data are of little use.

Additionally, dense spectra can make it very difficult to assign a transition line to its corresponding lower energy level, without which it is impossible to correctly extrapolate spectra to temperatures beyond those used in the experiment where they were measured. These difficulties are further exacerbated in environments at high temperatures, because their molecular spectra become extremely dense, with billions of features. In these cases, it is extremely challenging to obtain absolute line strengths, to assign transitions to their energy states, and even harder to obtain a complete spectra within a frequency range. Completeness is

an often overlooked but crucial parameter when evaluating a line list, as even very weak lines can have noticeable consequences to the spectrum of a molecule, either by becoming stronger at higher temperatures, or by collectively having an impact through sheer quantity. Extremely dense spectra like that which are produced by small organic molecules have so many tightly packed weak lines that it becomes impossible to extract a baseline from their spectra - these are called weeds. When Gl 229B first started being studied, existing experimental line lists were far too incomplete to predict its atmospheric spectra. The high temperatures in the atmosphere led to many more populated lines than could be modelled with existing line lists, leaving modelled spectra with unnatural opacity terminations in crucial molecules like water and methane [123].

As a more modern example, Figure 2.6 demonstrates the effect of completeness on a methane spectrum.

Figure 2.6: Spectra of the T4.5 dwarf, as observed by the SpeX instrument on the 3-m NASA Infrared Telescope Facility [40, 137], and modelled using the VSTAR model with both the STDS line list [10] and the 10to10 line list [190], and also using the BT-Settl model of Allard et al [5]. Figure replicated from Yurchenko et al [191].



The VSTAR [10] spectra of the T 4.5 dwarf could not be replicated by the existing experimental line lists, which contain approximately 350 000 transition lines, despite these including every strong transition in the room temperature methane spectrum. When using ExoMol's 10to10 methane line list [190, 191], which has almost 10 billion lines, the observed spectra is much better replicated. In fact, it was not possible to adequately model the observed spectrum with less than 3 billion lines [191]. This demonstrates the impact of completeness, particularly at high temperatures, when lines that had negligible contribution to the overall opacity at room temperature become active. The example of the T4.5 dwarf spectrum mostly illustrates the importance of completeness within the observed frequency range. However, an even more common issue arising from using experimental data in astrophysical observations is the incorrect extrapolation of the line intensities from the temperature the molecular spectra was measured at to the temperature of the observed body. A lot of molecular line lists used to model high temperature spectra are actually measured at room temperature and this can cause discrepancies between predictions and observations. This is effectively also a completeness issue; at room temperature, transitions due to high rotational excitations and vibrationally excited bands are not strong enough to be measured.

If a series of transitions have been measured at a specific temperature then, to extrapolate their intensities to represent an alternative temperature, one must understand the relationship between a line's intensity and its strength. A temperature dependent line intensity for a transition between initial and final energy states, E_i and E_f, can be calculated from a given temperature independent line strength with the following relationship:

$$I_{if}(T) = S_{if}\frac{e^{-E_i/k_bT}}{Q(T)}\left[1-\exp\left(\frac{-hc\nu_{if}}{k_bT}\right)\right]\frac{8\pi^3 N_A \nu_{if}}{(4\pi\epsilon_0)3hc} \tag{2.6}$$

where N_A is avogadro's number, ϵ_0 is the emissivity coefficient of the medium, ν_{if} is the measured transition frequency ($E_f - E_i$), and all other terms as defined in equation 2.4.

Then, the intensity of the same transition at a particular temperature, T_2, can be extrapolated from its measured temperature, T_1. This is done using equation 2.6 by considering the ratio of the intensities at the two temperatures:

$$I(T_2) = \frac{Q(T_1)}{Q(T_2)}e^{E_i/k_b(T_1-T_2)}\frac{\left[1-\exp\left(\frac{-hc\nu_{if}}{k_bT_2}\right)\right]}{\left[1-\exp\left(\frac{-hc\nu_{if}}{k_bT_1}\right)\right]} \tag{2.7}$$

It is clear that a correct temperature dependent-partition function, $Q(T)$, is required, although this is almost impossible to obtain experimentally (see section 4). Most importantly, unless a transition with measured intensity I_{if} can be assigned to its lower energy state E_i, this temperature extrapolation cannot be accurately performed. The difficulty behind assigning lines is further increased by the mixing between the upper vibrational states, making unambiguous labelling of transitions impossible [103].

It is, of course, very hard to resolve fine spectral structures and even more impractical to expect experimental line lists to include billions of assigned lines. Experimental methods, although extremely accurate, struggle to provide the quantity, quality and variety of spectral information required to correctly characterise astrophysical bodies, particularly at high temperatures. On the other hand, a theoretical approach rarely yields sufficiently accurate line lists without help from experimental fitting and refinement. Additionally, many molecules are still too complex to model from first principles, while retrieving their spectra experimentally is relatively straightforward. Most accurate data on molecular geometric structure is also still provided by analysis of laboratory spectra, particularly that of pure rotational spectra [160]. Ideally, public molecular databases should improve their data by incorporating assignments and intensities from theoretical models, while theoretical databases should continue to improve their *ab initio* models while validating and tuning their results with reliable experimental data.

The remainder of this chapter will provide an overview of the quantum mechanical approaches to creating a theoretical line list, using variational methods. The subsequent chapters will explain how a pure *ab initio* model is refined and validated by the existing experimental data, to ultimately produce a reliable, comprehensive phosphine line list.

2.3 Theoretical Line Lists

The current section is not just the result of extensive independent literature research but also the result of a compilation of knowledge acquired throughout my undergraduate, masters and current doctorate education, and through discussions with my colleagues, supervisors and collaborators. Throughout the remainder of this chapter, references will be included for specific sources of information, but a few textbooks were crucial to the overall understanding of the topic of theoretical line lists. These include the “Handbook of High-resolution Spec-

troscopy", specifically the chapter "High Accuracy Rotation-Vibration Calculations on Small Molecules" authored by Jonathan Tennyson [161], "Molecular Symmetry and Spectroscopy" by P. R. Bunker and P. Jensen [25], "Introduction to quantum mechanics" by D. J. Griffiths and E. G. Harris [61], and "Astronomical Spectroscopy", by Jonathan Tennyson [160].

A molecular line list consists of a description of energy states, and of the temperature dependent transitions between them. Diagram 2.7 provides an overview of the steps involved in producing the components of a line list and its subsequent spectra. The creation of a

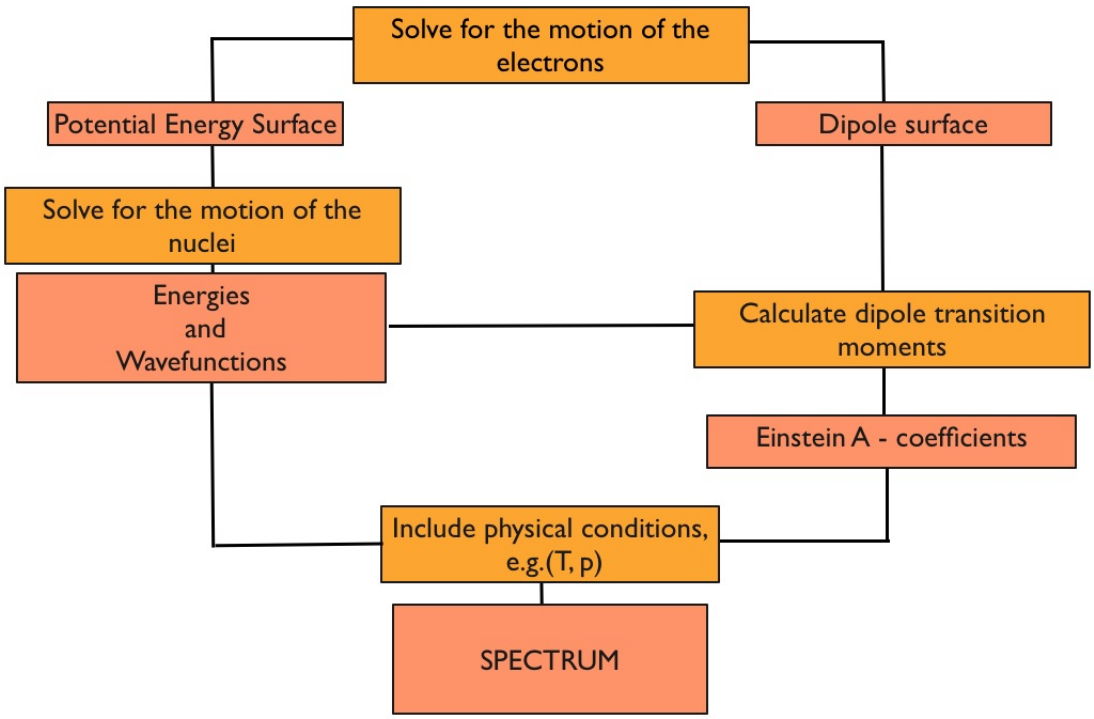


Figure 2.7: Creating a line list from first principles.

theoretical line list requires highly accurate potential energy surfaces (PESs), dipole moment surfaces (DMSs) and an ability to generate precise wave functions and eigenvalues for the nuclear Schödinger equation and the associated transition intensities.

The potential energy surface is a major source of error and, in practice, it is necessary to incorporate a certain amount of experimental refinement in its creation. Additionally, the accuracy with which high ro-vibrational states can be computed depends heavily on the size of the Hamiltonian matrix, as well as the expansions of the potential and kinetic energy operators. Once the energy levels and Einstein-A coefficients are computed, adequate knowledge of the partition function is necessary to establish the correct temperature dependence of spectral lines and its intensity.

2.3.1 Energy Levels

The production of all the components of a line list is a lengthy and computationally demanding process. The first bottleneck comes from solving the eigenvalue problem to obtain the energy levels of a molecule. This is a many step process that begins with solving the Schrödinger Equation for the motion of a molecule.

Solving the Schrödinger Equation

The Hamiltonian corresponding to the Schrödinger equation for a molecule with N atoms, each with charge Z_I, mass M_I and n electrons will have $3(N+n)$ degrees of freedom and can be written as:

$$\hat{H} = -\sum_{I=1}^{N} \frac{\hbar^2}{2M_I}\nabla_I^2 - \frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\nabla_i^2 - \sum_{I=1}^{N}\sum_{i=1}^{n}\frac{eZ_I}{|R_I - r_i|} + \sum_{I=1}^{N}\sum_{J<I}\frac{Z_I Z_J}{|R_I - R_J} + \sum_{i=1}^{n}\sum_{j<i}\frac{e^2}{|r_i - r_j|}$$
$$= \hat{T}_N + \hat{T}_e + V_{Ne} + V_{NN} + V_{ee} \tag{2.8}$$

where $\hat{T}_N$ is the kinetic energy operator for the motion of the nucleus, $\hat{T}_e$ is the kinetic energy operator for the electrons, V_{Ne} is the attractive potential between the nuclei and the electrons, V_{NN} is the repulsive potential between the nuclei and V_{ee} is the repulsive potential between the electrons. R_I and r_i are, respectively, the coordinates of the nuclei and the electrons, relative to a given origin. The kinetic and potential energy corresponding to the motion of the nucleus need not be derived analytically, and can be approximated numerically. The kinetic energy relates to the moments and products of the inertia of the molecule, so an appropriate choice of coordinate systems is important to facilitate its calculation.

The potential of a molecule is thus given by the various Coulomb interactions between the electrons and the nuclei of the molecule, which are only considered individually because of the Born-Oppenheimer approximation [20].

The Born-Oppenheimer Approximation

Molecular structure and subsequent spectra depend on many factors. The nuclei of the different atoms are not static, but rotate and vibrate. The electrons around them have a much smaller mass so they move comparatively fast. Then one can consider that any movement of

the nuclei leads to an instantaneous change of configuration for the electrons. This assumption leads to the Born-Oppenheimer (B-O) approximation, where the electronic and nuclear motions are treated separately. The overview of the calculation of a line list in Figure 2.7 implicitly depicts the B-O approximation by solving for the motion of the electrons separately from the nuclear motion calculation.

This treatment consists of representing the total wavefunction of the system, Ψ, as a product of the electronic and and nuclear wavefunctions, Ψ^{elec} and Ψ^{nuc}.

$$\mathbf{\Psi} = \Psi^{elec}(r_i; R_I)\Psi^{nuc}(R_I) \tag{2.9}$$

The nuclear wavefunctions depend only on the nuclear coordinates, R_I, and the electronic wavefunctions depend explicitly on the electronic coordinates, r_i, and parametrically on the nuclear coordinates, R_I [161]. This means that under the B-O approximation, a different wavefunction for the motion of the electrons can be obtained for each configuration of the nuclei. The Pauli principle, which stops two electrons from occupying the same space, is the biggest contributor to the repulsion forces between atomic nuclei separated by a small distance, R. Conversely, at very large separations, $R \to \infty$, the molecule dissociates. In between these extremes, the atoms are bound whenever the total molecular energy is larger than the sum of the individual atomic energies [160]. In particles where two electronic potential curves approach each other, the B-O fails because the nuclear and electronic motion can no longer be separated.

Within the B-O approximation, the electronic wavefunction satisfies the following Schrödinger equation

$$(\hat{T}_e + V_{Ne} + V_{ee} - E^{elec})\Psi^{elec} = 0 \tag{2.10}$$

Solving this equation yields the electronic energy eigenvalue E^{elec} corresponding to each of the electronic wavefunctions. E^{elec}, together with the inter-nuclear potential, V_{NN}, provides the potential energy surface (PES) upon which the nuclei move, $V(Q)$:

$$V(Q) = E^{elec}(Q) + V_{NN}(Q) \tag{2.11}$$

where Q are the internal coordinates.

The Potential Energy Surface

It is the B-O approximation that gives rise to the PES, which will be one-dimensional for a diatomic molecule, otherwise it will have dimensions $3N - 6$, given a molecule with N atoms. Due to the parametric dependence of the electronic wavefunction on relative nuclear coordinates, the PES depends on the internal coordinates, Q. It is, however, independent of the mass of the nuclei, and so conserved for all isotopologues of a molecule. The PES will have dimensions of $3N - 6$ for a molecule with N atoms. The B-O approximation also neglects the effect of the nuclear kinetic energy operator, $\hat{T}_{nuc}$, on the electronic wavefunction, Ψ^{elec}; this can be corrected by adding a term accounting for this interaction, $\langle \Psi^{elec} | \hat{T}_{nuc} | \Psi^{elec} \rangle$, to the potential energy function. Additional corrections to the PES will have a smaller impact, but include relativistic corrections and Lamb shift* [160].

The total energy for a particular state of the system, E, can then be obtained by using the potential, $V(Q)$, to solve the Schrödinger Equation for the nuclear motion:

$$(\hat{T}_n + V(Q) - E)\Psi^{nuc} = 0 \tag{2.12}$$

The PES has a semi-parabolic shape due to its connection to simple harmonic motion, which means it has a lowest energy corresponding to its minimum. This is called the zero point energy, or ZPE, and cannot be equal to zero, as that would violate Heisenberg's Uncertainty Principle; if a state exists that allows for the system to remain motionless, then it would be possible to determine both its position, x, and its momentum p, which is forbidden under the Uncertainty Principle: $\sigma_x \sigma_p \leq \hbar/2$.

The attraction between nuclei at large internuclear separations is weaker than their repulsion at short bond lengths; this is what distorts the parabolic shape of the PES and gives it an asymmetric shape with a dissociation limit, D_e. This potential would otherwise be able to hold an infinite number of energy levels, but its actual shape means that the levels effectively end at dissociation. For stable molecules, the potential surfaces that arise from solving the electronic Schrödinger equation have two important features: an asymptotic dissociation energy and a minimum, corresponding to the equilibrium bond length, R_e, and the associated zero point energy of the molecule. Figure 2.8 demonstrates the difference between the potential for a harmonic oscillator and that of a real system, for a diatomic molecule.

*Lamb shift is a small different in energy for levels where the Dirac equation predicts the same energy; this is due to the interaction between the electron and the quantum electrodynamic vacuum.

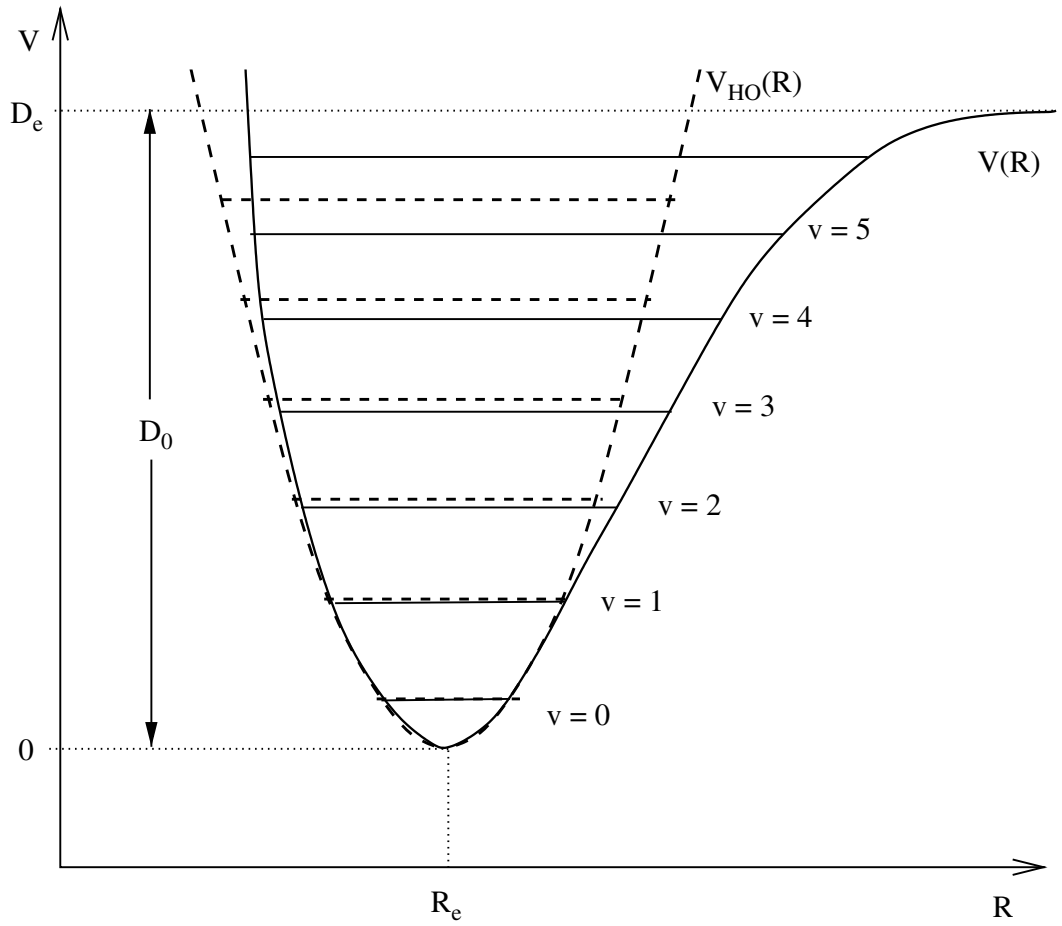


Figure 2.8: Potential curves for a diatomic molecule given internuclear separation, R. The harmonic oscillator approximation, V_{HO}(R), is shown in dashed lines and the more realistic potential, V(R), in full lines. Because the ZPE is non-zero, the dissociation energy, D_0, is actually smaller than the dissociation limit, D_e. Figure from "Astronomical Spectroscopy" [160].

The difficult dependence of the potential on the position of the nuclei makes it impossible to obtain an algebraic solution to $V(Q)$. The potential curve for a molecule resembles a parabola, and in a harmonic oscillator approximation it remains so. The true shape of this surface is very complex and must either be approximated or obtained numerically; a parabolic shape is usually a good approximation at small bond lengths but becomes increasingly unreliable at large distances, which correspond to higher energy levels. An harmonic oscillator would have an infinite number of evenly spaced energy levels. Conversely, the states in a more realistic potential energy curve will become increasingly dense and eventually reach disassociation. It is important to obtain the best possible shape of the PES because it is the shape, not the absolute energy of the surface, that determines the value of the ro-vibrational energies.

Vibrational and Rotational Wavefunctions

The total energy, E, includes the energy due to both the electric and nuclear motion. The energy associated with the former is much greater than that with the latter. Within nuclear motion, the energy corresponding to the vibrational motion is generally greater than that corresponding to the rotational motion of the nucleus. It is common to consider the motions of a molecule as divided into the internal motion of the molecule (vibrational), the rotation of the whole molecule and the translation of the molecule as it moves through space. The energy associated with the translational motion gives a continuum spectrum, which is unhelpful if one is trying to extract the discrete energy levels of the molecule. It is therefore necessary to isolate the three coordinates associated with this motion, which is possible as translation can be completely separated from the vibrational and rotational motions. The three coordinates which describe the translational motion are the laboratory coordinates of the molecule's centre of mass. The remaining vibrational and rotational motion coordinates have dimensions of $3N - 3$, and cannot be rigorously separated. These are defined as space-fixed coordinates. However, considering body-fixed coordinates (also known as molecule-fixed coordinates), where the vibrational and rotational motions are separated, provides a good first approximation:

$$\mathbf{\Psi_{nuc}}(Q) = \Psi_{vib}(R)\Psi_{rot}(\theta, \phi) \tag{2.13}$$

where the rotational motion is represented by the Euler angles, θ and ϕ, that define the orientation of the molecules with respect to the space-fixed and body-fixed coordinates, and the vibrational motion is represented by a set of internal coordinates, R, which, in practice, also depend on the rotational quantum numbers, described below. Together, these $3N - 6$ coordinates, Q, define the relative position of the nuclei. Internal coordinate choice is a vast subject and most theoretical calculations make use of multiple types, converting between them. In general, different coordinates are used at different stages of the calculation, and also depend on the desired outcome of the process, be it computational efficiency, versatility, or structural insight. This is further complicated by the fact that molecules with different number of atoms or configuration require their own specific internal coordinate system to describe their nuclear motion. Often, coordinate systems that simplify the description of the kinetic energy operator will describe the potential energy poorly, or vice-versa [193].

Vibrational and Rotational Basis Functions

The energy for which the vibrational motion is responsible, E_v, can be obtained by considering the Schrödinger equation:

$$\hat{H}_v|v\rangle = E_v|v\rangle \tag{2.14}$$

where $|v\rangle$ are the vibrational basis set functions, usually constructed as a product of standard analytical one-dimensional functions, $\phi_{v_{n_R}}(R)$, given by:

$$|v\rangle = \prod_{\nu_i} |v_\nu\rangle = \phi_{\nu_1}(R)\phi_{\nu_2}(R)...\phi_{\nu_{n_R}}(R) \tag{2.15}$$

where n_R are the number of internal coordinates chosen, and correspond to each vibrational mode, ν_n, of the molecule. Most molecules will have $3N-6$ vibrational modes, except linear molecules which will have $3N-5$ vibrational modes, since they cannot rotate about their molecular axis. The basis functions can also be obtained numerically by solving a reduced one dimensional Schrödinger equation for the number of degrees of freedom of each molecule.

The equivalent rotational basis functions can be constructed from the rigid-rotor functions $|J, K, m, \tau\rangle$, where J is the rotational angular momentum quantum number, K is the projection of the angular momentum on the body-fixed axis z, m is the space-fixed projection quantum number and $\tau = 0, 1$ indicates the rotational parity $(-1)^\tau$ as defined in Yurchenko et al. [189]. The K quantum number follows the rule $-J \leq K \leq +J$. Molecules will have different sets of rigid-rotor functions depending on their structure.

Ro-vibrational Hamiltonian

Considering the vibrational and the rotational basis functions as above, the matrix elements of the ro-vibrational Hamiltonian, $\hat{H}_{rv}$, will then be calculated using the basis functions:

$$|v, J, K, m, \tau\rangle = \prod_{\nu_i} |v_\nu\rangle \times |J, K, m, \tau\rangle \tag{2.16}$$

The rotation-vibration energies and wavefunctions can then be obtained by performing a standard variational calculation. This requires that the ro-vibrational Hamiltonian matrix, H_{rv}, be diagonalised. The basis set required to obtain an exact description of the Hamiltonian matrix would have to be infinitely large, so the number of basis functions used needs to be truncated to make the diagonalization of the Hamiltonian possible. Basis functions have

coordinates that are also chosen specifically to simplify the Hamiltonian as much as possible. Truncating the number of basis functions affects the resulting energy levels, as every additional basis function would increase their value slightly. It is therefore important to perform convergence tests with an increasing number of basis functions, until one can ensure that the values of calculated energy levels are no longer materially changing.

With the right choice of basis set, the Hamiltonian matrix can be given a block diagonal structure, where each block corresponds to independent J blocks, and each J block in turn has a block diagonal structure, where each block corresponds to one irreducible representation of the molecule's symmetry. The diagonalization of the Hamiltonian can therefore be simplified by being performed separately for each block, the size of which will depend on the rotational excitation, J, the choice of basis functions and the complexity of the molecule. To ensure this is computationally feasible, sufficient computer memory needs to be made available for both the matrix blocks and for their associated eigenvectors; this is not always possible with present day computer memory. Nonetheless, it can still be extremely difficult to solve for the eigenvalues of individual J blocks, with sizes easily reaching 200 000 x 200 000 for molecules with only four or five atoms. Even by separating the Hamiltonian matrix by symmetry and J blocks, matrices can still be far too large for diagonalization to be possible with current hardware. Making the diagonalization process efficient is the focus of various research areas, and usually requires highly efficient algorithms and state of the art computation. Until very recently, the RAM size and CPU power required to solve the eigenvalue problem for even small molecules simply did not exist.

Ro-vibrational Coupling

Equation 2.13 is a good approximation but the rotation and vibrational motion are nonetheless coupled, although fairly weakly. For example, the rotation of a symmetric top molecule causes vibrational modes to excite each other due to the effect of Coriolis forces [22]. This causes an interaction between the rotation and vibrational motion of the molecule, which can be addressed by either explicitly including a Coriolis operator in the Schrödinger equation, or by adding a Born-Oppenheimer correction to the equation for the potential (2.11). The latter can be done using expansion coefficients in the nuclear motion wavefunctions to account for the rotation-vibration interactions.

Energy Contributions

The combined rotational and vibrational energies obtained from separating the nuclear motions provides the ro-vibrational energy levels of the system, $E_{rv} = E_{rot} + E_{vib}$. As mentioned before, the rotational contribution to the total energy is considerably smaller than that of the vibrational motion. At least for the low-lying energy levels, a rough estimation of the ro-vibrational energy can also be made by using spectroscopic constants derived from perturbation theory, but these approximations (Harmonic Oscillator and Rigid Rotor) deviate from the actual values for the energy levels at higher excitations [160].

If the molecule can only be in its electronic ground state, then E_{elec} will be a constant, and the total energy of a particular state, E can be given by:

$$E_{total} = E_{elec} + E_{rot} + E_{vib} \tag{2.17}$$

where the energies here can have units of frequency (Hz), wavelength (μm) or, commonly in molecular spectroscopy, wavenumbers, $\tilde{\nu}$, where:

$$\tilde{\nu}[\mathrm{cm}^{-1}] = \frac{10^4}{\lambda[\mu\mathrm{m}]} \tag{2.18}$$

Labelling convention for molecular transitions is to label a lower energy state E'' and an upper energy state E', so that absorption is $E' \leftarrow E''$ and $E' \rightarrow E''$.

It is worth noting that each motion has a very different energy contribution, as demonstrated in figure 2.9. Vibrational excitations increase in energy evenly, while the energy steps due to the rotational motions widen with J. Consequentially, the large spectral structures due to the vibrational motion contain fine structures from simultaneous rotational motion.

The rotational energy of a state can be roughly estimated by the rigid-rotor approximation, $E_{rot} = B_0 J(J+1)$, where B_0 is a molecule's rotational constant. Similarly, the vibrational component of the energy for a molecules with M vibrational modes can be estimated by the harmonic oscillator approximation, with $E_{vib} = \hbar \sum_{i=1}^{M} w_i(\nu_i + \frac{1}{2})$, where w_i is the frequency and ν_i is the vibrational quantum number.

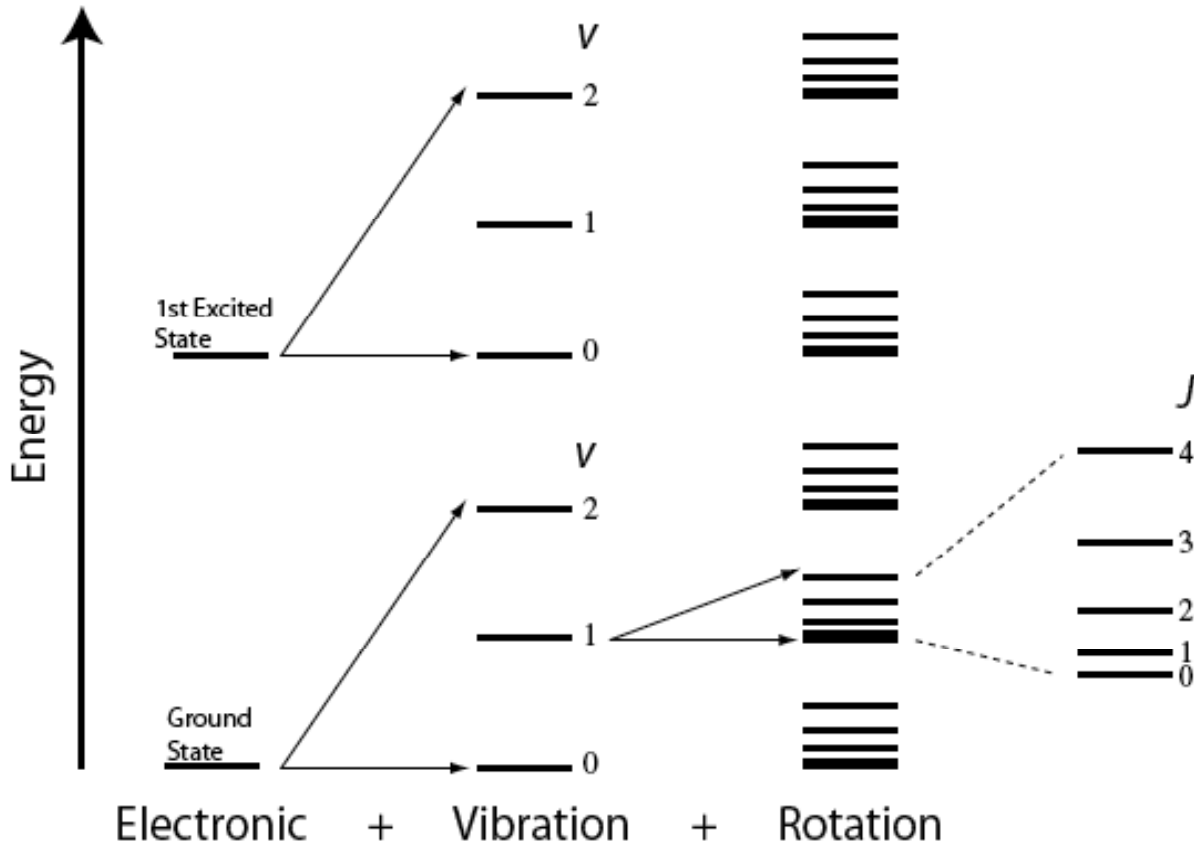


Figure 2.9: Hierarchy of energy levels.

Symmetry

The molecular Schrödinger equation is solved by diagonalising matrix representations of the Hamiltonian; these calculations are facilitated by symmetry. Molecular symmetry operations are defined as operations that keep the wavefunctions, and consequently the molecular energy, unchanged; a symmetric operation is one that commutes with the Hamiltonian. Symmetry imposes selection rules on molecular transitions through the vanishing integral rule:

$$\int \Psi_i^* \hat{H} \Psi_f d\tau = 0 \tag{2.19}$$

unless the integrand, $\Psi_i^* \hat{H} \Psi_f d\tau$, contains a totally symmetric component in the symmetry group of the Hamiltonian. Effectively, elements of the Hamiltonian matrix are only non-zero if they are the results of basis function combinations from the same symmetry. This factorisation of the Hamiltonian matrix drastically reduces the diagonalization process, as the different symmetries form independent blocks around the diagonal, leaving the remaining matrix elements as zero. These blocks are organised as irreducible representations of each

of the molecular symmetries in a molecule. Symmetry can be incorporated into the nuclear motion calculations by choosing wavefunctions that transform according to one-dimensional representations.

Symmetry imposes limitations on the allowed transitions in a molecule. For all molecules, rotational symmetry obeys the strict conditions:

$$\Delta J = J' - J'' = 0, \pm 1 \text{ and } \mathrm{J}' + \mathrm{J}'' \neq 0 \tag{2.20}$$

Further selection rules depend on the individual molecule's symmetry point group.

Fitting the PES

Since the calculated *ab initio* PES is approximated using a finite grid of basis functions, it requires fitting to accurately represent the true PES. This is usually done by fitting the electronic energies to an appropriate analytical polynomial function relating to each of the coordinates used to calculate the surface. This function must be flexible as it requires a large number of often correlated parameters, and the shape of the PES when away from equilibrium is very sensitive to small changes to its original functions. If reliable experimental data on energy levels exists, then it can be incorporated in the fitting process to improve the shape of the *ab initio* PES to better match the measured energies. Effectively, this process requires multiple interactions where the PES is used to calculate ro-vibrational energies, then these are compared to their corresponding experimental values; the parameters that define the PES are then adjusted to slightly modify its shape and consequently improve the agreement between the calculated and the measured energies. Due to the sensitivity of the PES, only approximately 10% of the parameters need any significant refinement for the overall shape of the PES to show considerable improvements. Surfaces that undergo this empirical adjustment are usually referred to as a spectroscopically determined potential energy surfaces.

In summary, the B-O approximation separates the electronic motion from the nuclear motion. The electronic Schrödinger equation can then be solved for the situation when the nuclei are considered to be fixed in space, where the eigenvalues will be the potential surface for each electronic state, $V(Q)$. The nuclear motion Hamiltonian will now be the kinetic energy of the molecule together with the potential obtained from the electronic solution of the Schrödinger equation. Solving the nuclear Schrödinger equation will then yield the energy eigenvalues. These will each correspond to a nuclear-motion wavefunction, composed of

a symmetry allowed combination of basis functions; this is how each energy level can be assigned to a set of quantum numbers. Now that the energy levels are computed and labelled, all that there is left to obtain is the transition intensities.

2.3.2 Transition Intensities

As mentioned before, the diagonalization of the matrix Hamiltonian to obtain the energy eigenfunctions is the first bottleneck in the computation of a line list. The second is the calculation of the transition line strengths required to obtain the transition intensities. Even for reasonably simple molecules, like phosphine (PH_3) or methane (CH_4), a complete line list can contain many billions of lines [190] representing transitions between several million individual energy levels. Transitions between states having a single J pair can add up to over half a million lines. Fortunately, algorithms for the calculation of line strengths allow for higher granularity and consequently higher levels of parallelism than those used for the calculation of the eigenvalues. Most high performance computing clusters have hundreds of computer nodes, each containing multiple core processors, so the speed of this stage of the line list calculation is effectively only limited by the number of processors that can be made available. Graphics processing units, or GPUs, are cheap and power efficient processing units that have a highly parallel structure and are therefore very useful for the purpose of calculating transition intensities. GPUs have been used extensively throughout this project [3].

The intensity of a transition that occurs between two energy states will be dependent on each state's probability of being occupied, as stipulated by the Boltzmann distribution, N$= \exp(-E/k_bT)$; the population, N, of an energy level is therefore temperature dependent. In fact, given how densely packed the energy levels of typical molecules are, they can act as a thermometer for their local environment [160]. At very low temperatures ($\leq$20 K), only few energy levels will be populated, mostly in rotational energy steps within the vibrational ground state. At higher temperatures, vibrational energy separations start to be relevant. For most molecules, the temperature at which the first excited electronic state would become occupied is so high that the molecule dissociates before it can reach it, and as such it can only be found in its electronic ground state.

The intensity will vary with temperature, as this affects the population of different energy states, but a temperature independent transition strength can be calculated by considering a molecule's electric dipole.

Dipole Moment

To calculate the probability for a transition between state i and f, a transition dipole, μ^{if}, is needed. A molecule's dipole is the result of the arrangement of the charges of its constituents. If the overall charge distribution of a molecule is asymmetric, then it has a permanent dipole moment. Otherwise, a dipole is only acquired when it is induced by the vibrational motion of the molecule. Distortions from the equilibrium configuration of a molecule without a permanent dipole can give rise to an instantaneous dipole moment. For a molecule to exhibit purely rotational transitions (without vibrational excitations), it must possess a permanent dipole moment. For example, CO_2 has only ro-vibrational spectra because a dipole moment is only induced when it oscillates. A vibrational excitation is "dipole-allowed" if the derivative of the dipole moment (permanent or induced) is non-zero; all polyatomic molecules have at least a few vibrational modes where this occurs and so can exhibit a ro-vibrational spectra. The transition dipole, μ^{if}, between two states can be calculated as the expectation value of a dipole operator and can be calculated by integrating over all nuclear degrees of freedom using a dipole moment, $\mu(Q)$:

$$\mu^{if} = \langle \Psi^i | \mu(Q) | \Psi^f \rangle \tag{2.21}$$

where Ψ^i and Ψ^f are the nuclear motion wavefunctions for the initial and final energy states, respectively, and the dipole moment, $\mu(Q)$, is the instantaneous dipole moment of the molecule at a given coordinate configuration, Q.

The dipole moment describes how the positive and negative charges of the particles in a molecule are distributed. The dipole moment is the functional form of the dipole moment surface (DMS), which can be automatically obtained for a grid of geometries when calculating the electronic wave function. As a consequence, the accuracy of the dipole moment surface, and subsequent transition probabilities, is dependent on the quality of the potential energy surface used to generate the wavefunctions for the initial and final states.

Unlike PES calculations, which usually require some experimental fitting to obtain a sufficiently accurate shape, theoretical DMSs are extremely reliable, and often produce much more accurate line intensity predictions than those obtained for the energies through the PES. This is helped by the fact that DMS calculations converge much faster with basis set size than their potential counterparts. Nonetheless, the DMS can still be improved by adding corrections to the *ab initio* surface, including B-O diagonal corrections, relativistic effects,

spin-orbit coupling and core correlation. It is worth noting that the enhancement granted by these corrections is not always evident; in water, for example, the corrections due to core correlation and relativistic effects effectively canceled each other out [95]. Similarly to the PES, obtaining a high accuracy description of the DMS is much easier near equilibrium than maintaining a smooth and accurate curvature for determining intensities at higher vibrational excitations.

The transition dipole can be represented as the derivative of the dipole moment function of a molecule, where single excitations from the vibrational ground state can be considered as a result of the first derivative, while transitions to states with m quanta of excitation will correspond to the m^{th} derivative of the electronic energy. This means that even minute irregularities in the shape of the DMS will be evident when calculation transition dipoles for high vibrational excitations. It is therefore important to ensure that the DMS is smooth and behaves appropriately, which is particularly challenging when a molecule approaches dissociation. Additionally, while it is theoretically possible to calculate the dipole moment for every grid point used to generate the wavefunctions of the molecule, this is only manageable for small molecules. It is common to instead create a functional form of the DMS to interpolate between calculated points, fitted to agree with the expected behaviour of the DMS [162].

Nuclear spin and statistical weight

The intensity of a transition is also proportional to the statistical weight of its initial level, which is related to its J value by a factor of $(2J + 1)$ and to the total degeneracy of its state. Statistical weights are temperature independent factors of the Einstein-A coefficients but they are symmetry dependent. Symmetries with the lowest weight correspond to "para" states while symmetries with the highest weight correspond to "ortho" states. The statistical weight of different symmetries define their relative intensity in a spectrum.

The degeneracy of a state depends on the change of symmetry of its wavefunction when any two identical nuclei are exchanged. The Pauli Principle requires that the resulting wavefunction to be either unchanged (even symmetry) or antisymmetric (odd symmetry), which occurs, respectively, when the nuclei are bosons (integer spin) or fermions (half integer spin). The symmetry of the total wavefunction, Ψ_{Total}, depends on the symmetry of its components:

$$\Psi^{Total} = \Psi^{elec}\Psi^{vib}\Psi^{rot}\Psi^{ns} \tag{2.22}$$

where Ψ^{elec}, Ψ^{vib}, Ψ^{rot} and Ψ^{ns} are the electronic, vibrational, rotational and nuclear spin wavefunctions, respectively; these will themselves be even or odd for different J values. All allowed total wavefunctions will have either even or odd symmetry, depending on the symmetries of each of its components. However, due to the possible degeneracy of the nuclear spin functions, there can be several times more eigenstates of one symmetry, which will result in a degeneracy factor that increases the line strength of the transitions occurring from energy states corresponding to those symmetries. This factor is called the nuclear spin degeneracy, g_{ns}, and each molecular point group will have a g_{ns} value for each of its allowed symmetries. The nuclear spin degeneracy is mostly relevant as a statistical weight, although it also has a small effect on the molecular Hamiltonian as a hyperfine term which makes extremely small contributions to the energy values.

Nuclear statistical weights are further described for phosphine and ammonia in chapter 4.1.

Line Strengths

The ro-vibrational wavefunctions obtained from solving the Schrödinger equation for the motion of the nucleus can be combined with the dipole moment to calculate the line strength of a transition between initial and final energy states (eigenfunctions Ψ^i and Ψ^f, respectively):

$$S_{if} = \sum_{A=X,Y,Z} |\langle \Psi^f | \bar{\mu_A} | \Psi^i \rangle|^2 \tag{2.23}$$

where μ_A is the electronically averaged component of the dipole moment along the space-fixed axis $A = X, Y, Z$ (see also [192]). Due to the vanishing integral rule, allowed transitions must not alter the nuclear spin state of the system. This leads to a set of rigorous selection rules with respect to molecular symmetry, which vary between molecules. For all molecules, line strengths can be calculated in isolated collections of J pairs, due to the strict rotational rules as defined by equation 2.20. If all rotational excitation (J value) eigenvalues are available and the required number of processing cores is available, all J pair calculations could theoretically be performed simultaneously.

If the line strength value for each transition were to be obtained directly from equation 2.23, it would be necessary to perform $3N^2$ operations, where N is the size of the basis set. Considering that a line lists can have billions of transitions and, for large values of J, N can reach 10^6, this is not feasible. Instead the evaluation of the integral can be done only

for effectively non-zero elements, which drastically reduces the size of the problem. By neglecting any elements that do not contribute to the final line strength, large, but sparse eigenvectors become much easier to handle. Nonetheless, many molecules have hundreds of thousands of energy states corresponding to a single J value with dense eigenvectors, so transitions from and to these states alone can take months to compute.

Once the line strength of a transition is calculated, it can easily be used to obtain the Einstein-A coefficient or the transition intensity as described in equations 2.5 and 2.4.

It is the quality of the DMS that is crucial for accurate calculations of line strengths. Nonetheless, the accuracy of the line strengths is still dependent on the underlying PES. This is because although the transition strengths and the energy levels are calculated separately, the PES determines the quality of the ro-vibrational wavefunctions, and consequently has an impact on the accuracy of the line strengths. This underlying responsibility of the PES is particularly important to avoid mixing between wavefunctions; when states with similar energy values happen to have the same rigorous quantum numbers, interactions between them can lead to intensity "stealing", where an otherwise weak transition can become artificially stronger by many orders of magnitude. Having an accurate PES mitigates this effect by ensuring these neighbouring energy states are well defined and distinguishable.

2.3.3 The Complete Spectrum

The temperature dependent spectrum of a molecule, before broadening, is a collection of zero-width transition lines at the transition frequency $\nu_{if} = E_f - E_i$, with a line intensity height corresponding to the temperature dependent probability of that transition occurring.

At low frequencies, or long wavelengths, is where pure rotational spectra lie. Molecules can only have pure rotational transitions if they have a permanent dipole moment, and the intensities of these rotational transitions grow with the square of their molecules' dipole moment. In cold environments, like the interstellar medium (ISM), molecules cannot gain sufficient energy to exhibit anything but a rotational spectra, but as temperature increases, so does the incidence of vibrational excitations. As frequency increases towards the mid-infrared, vibrational transitions become relevant and this is referred to as the rotation-vibrational, or ro-vibrational, spectra. Ro-vibrational transitions become particularly significant in hot environments such as warm exoplanets, hot jupiters, brown dwarfs and other cool stars.

Rotation-vibrational spectra is structured by bands, which are dependent on the vibrational quantum numbers of the energies levels between which the transitions occur. There are four main types of vibrational bands: fundamentals (between the vibrational ground state and the first vibrational excitation of each mode), overtones (between the vibrational ground state and vibrational states above the first excitation), combination bands (between the vibrational ground state and a combination of excitations of different modes) and "hot" bands (between excited states). The bands for transitions between excited states are called "hot" bands because at room temperature only the ground state is highly populated; it is only when the temperature increases that vibration excitations are significantly further excited. Usually the fundamental bands are the strongest in intensity.

A high resolution spectra will show vibrational bands with fine rotational structures corresponding to rotational and vibrational excitations happening simultaneously. Within a vibrational band, transitions with rotational quantum number changes of $\Delta J = +1$ fall on the short wavelength side of the band centre, which is called the R-branch of the band. Similarly, transitions with $\Delta J = -1$ fall on the long wavelength side of the band centre, called the P-branch. The R and P branch transitions are evenly spaced in approximately $2B_0$ steps. If a molecule has allowed transitions with no change in J ($\Delta J = 0$), then they form the Q-branch. These all fall roughly at the band origin, giving the spectra a sharp peak at the centre (see figure 2.10).

The rotational branches can usually still be noticeable in vibrationally excited bands. However, bands will overlap, and ro-vibrational spectra can become incredibly rich, even for molecules with very few atoms, as soon as temperatures become elevated. The density of possibly occupied energy levels becomes so high that even within a small frequency range of a few hundred wavenumbers, billions of individual transitions can be present.

In molecules that survive excitations beyond their electronic ground state, transitions between electronic levels mostly occur in the ultraviolet spectral region.

Labelling the states

In its simplest form, a transition line can be described by the initial energy level, E_i, the transition frequency, $\tilde{\nu}_{if}$, and the temperature independent transition line strength, S_{if}. Alternatively, the frequency can be omitted, as long as the final energy level, E_f, is provided. This information is sufficient to obtain an overview of the spectra, but to understand it well

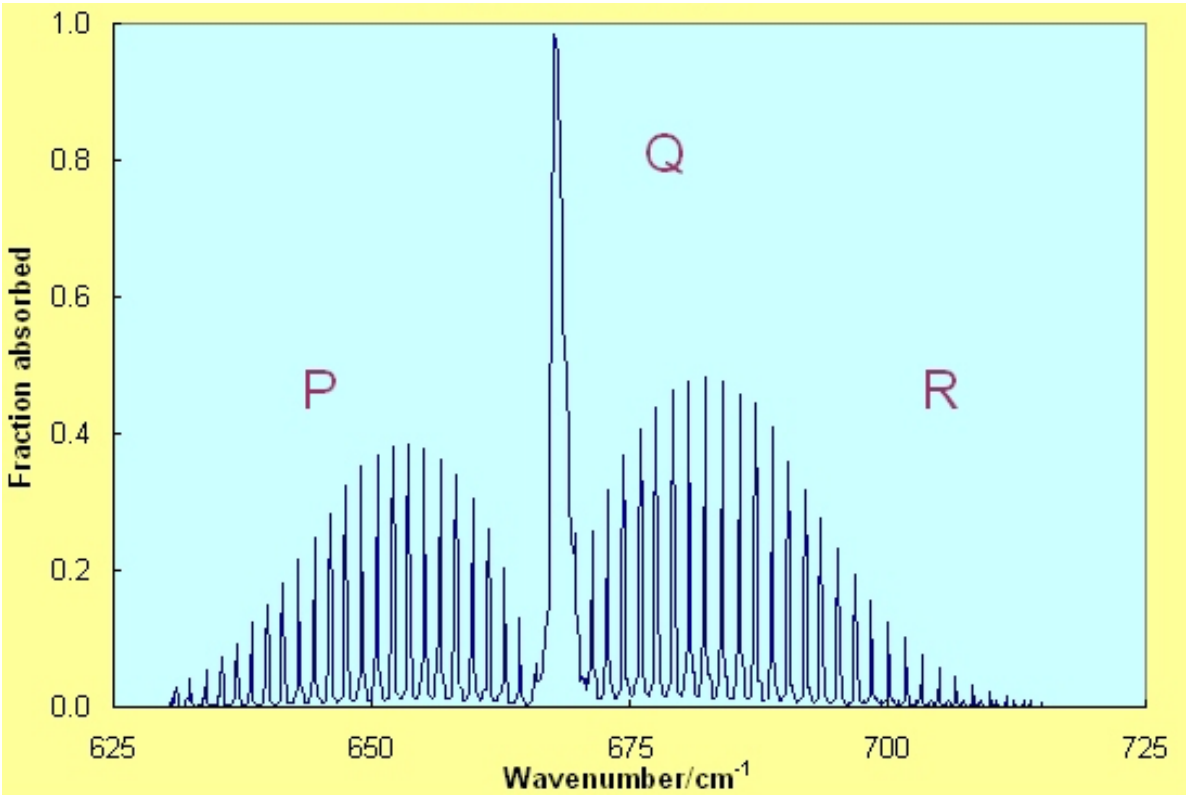


Figure 2.10: Rotational spectra at the fundamental vibrational band of CO_2, highlighting the P, Q and R branches. Image taken from Barrett Bellamy Climate Spectral Transitions.

enough to compare it to experimental data, and to create appropriate cross-sections, it is necessary to also have a description of the energy levels in terms of its symmetry and quantum numbers. In experimental spectroscopy, assigning energy levels with appropriate quantum labels can be a very arduous process, particularly for dense spectra. Theoretical spectra, on the other hand, produces energy levels with associated quantum numbers, as an inherent part of the calculation of the energy eigenvalues. The rigorous quantum numbers (J and symmetry) are automatically assigned, while the remaining quantum numbers are usually only approximately distributed. The usual minimum number of labels required to describe an energy state is the energy value, the degeneracy of the state and the J number. It is also common to label its total symmetry, as well as the K (projection of the angular momentum) quantum number and the vibrational quantum numbers representing the levels of excitation for each of the vibrational modes. Since molecules typically have a large number of energy levels, these quantum numbers are rarely sufficient for an unambiguous labelling of every energy state of a molecule, so additional quantum descriptors are usually provided. Nonetheless, it becomes increasingly impractical to assign unambiguous sets of quantum labels to each energy level

when there are several million of them, which is the case for a typical molecule.

The labelling choices made by different sources of spectroscopic data very wildly. This makes comparison of energy levels, and consequent comparison of transition frequency, between databases very difficult, and is made even harder due to the fact that only the total symmetry and J are strictly good, or rigorous, quantum numbers. The challenges surrounding the assignment of quantum numbers, and the comparison of transitions lines between theory and multiple experimental databases are addressed in chapters 3 and 5 in detail for the phosphine molecule. The ExoMol perspective on energy state labelling is also discussed at length in Tennyson *et al.* [164].

Chapter 3

Computing a Preliminary Phosphine Line List

The aim of this work is to develop a comprehensive line list for phosphine. At the beginning of this project, knowledge of phosphine was limited to a few theoretical and experimental sources. By the time the work presented here was completed, a few additional experimental sources of phosphine data had appeared. Collectively, these experimental sources, although accurate, are still limited in multiple ways. In particular, all the existing experimental data is appropriate only for use at low temperatures.

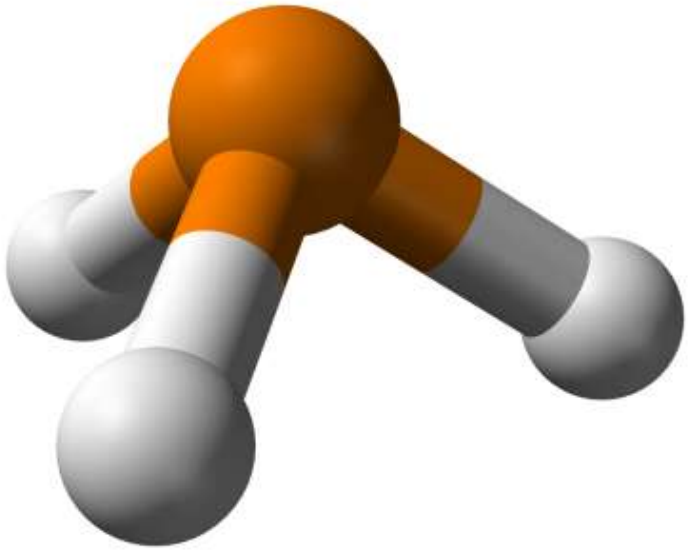
In order to establish and test the production of a final line list for phosphine, complete to high temperatures ($\leq$ 1500 K), it is first pertinent to perform a simulation of the room-temperature spectrum for this molecule, which can be tested against existing reliable experimental sources. This theoretical room temperature version of the line list is described in this thesis as SYT (Sousa-Silva, Yurchenko and Tennyson) and sometimes referred to as the 'cold' or 'cool' line list. The cool phosphine line list can be considered as the precursor to the final line list, SAlTY (pronounced 'salty'), and is the phosphorous analogue of the ammonia line list computed by Yurchenko *et al* [185], using the same program and similar methods.

In this chapter the existing work on phosphine spectra and associated data is described, then the simulation of phosphine is outlined, followed by a comparison of the room temperature line list to the best of the experimental data available.

3.1 Phosphine Characteristics

Phosphine is a semi-rigid, relatively stable, and oblate, symmetric-top molecule belonging to the group-15 hydrides. It is one of the lightest pyramidal molecules and, as such, any features that are inherent to molecules of this type will be particularly pronounced in the spectra of PH_3. The phosphorous atom is positioned on the axis of symmetry, perpendicular to the plane containing the equidistant three hydrogen atoms (see figure 3.1). As it is common with molecules having C_{3v}(M) symmetry [25], there is a splitting of rotational levels with $K = 3n$ (where $n \geq 1$) in non degenerate vibrational states [29].

Figure 3.1: Model of the phosphine molecule, with the phosphorous atom in orange and the three hydrogens in white. Bond lengths and angles from Greenwood & Earnshaw [60]. Image generated in Accelrys DS Visualizer [174].



Recorded phosphine spectra show no inversion splitting. The tunnelling effect found in ammonia is expected to occur in phosphine but so far attempts to detect it have failed due to its considerably higher potential energy barrier (12300 cm^{-1}) between the two symmetry-related minima [17, 148]. The phosphine splitting is further addressed in section 6.3. In the absence of the inversion splitting, phosphine is characterized by the C_{3v}(M) molecular group symmetry, i.e. each eigenfunction transforms according to one of the three irreducible representations A_1, A_2 and E, the latter of which is two-fold degenerate. Radiative and collisional transitions between A and E symmetries are absolutely forbidden.

The PH_3 molecule has four atoms and as such four vibrational modes. Phosphine has near degenerate ν_1 symmetric and ν_3 antisymmetric stretching modes with frequencies of about twice that of the symmetric and asymmetric bending modes, ν_2 and ν_4 [176]. It is therefore natural to consider its spectrum in terms of polyads, which will be described in section 3.4.

3.2 Previous Work on Phosphine

3.2.1 Experimental Work on Phosphine

Experimental observations of PH_3 showed that there are various interactions between its spectral bands. There is a strong Coriolis interaction between the ν_2 and ν_4 bending bands which causes large distortion effects in observations [158]. In fact, many bands are affected by several different types of interaction. As an example, figure 3.2 shows the interactions affecting the phosphine pentad, in the 1885 – 2478 cm^{-1} spectral window.

Figure 3.2: Interacting bands of the PH3 pentad [103].

	$2\nu_2$	$\nu_2+\nu_4{}^{\ell}{}_4$	$2\nu_4{}^{\ell}{}_4$	ν_1	$\nu_3{}^{\ell}{}_3$
$2\nu_2$	K-type Interaction Diag	Coriolis	Coriolis Fermi	Fermi	Coriolis
$\nu_2+\nu_4{}^{\ell}{}_4$		ℓ-type interaction Diag	Coriolis Fermi	Coriolis	Fermi
$2\nu_4{}^{\ell}{}_4$			ℓ-type interaction Diag	Coriolis Fermi	Coriolis Fermi
ν_1				K-type interaction Diag	Coriolis
$\nu_3{}^{\ell}{}_3$					ℓ-type interaction Diag

Coriolis refers to the rotation-vibration interactions between energies that are described in section 2.3.1 and Fermi refers to Fermi resonances, where bands with similar energies will split as a result of the mixing of their wavefunctions. With Fermi interactions, the affected bands (known as a 'Fermi doublet') partially exchange intensities and spread out in frequency. Both K-type and l-type interactions occur due to degeneracies in upper states with the same symmetry, for $\Delta K = \pm 3$ and $\Delta K = \Delta l$ [103]. K refers to the projection of the angular

momentum of the molecule on the z-axis and l describes the projections of the angular momenta corresponding to degenerate vibrational upper states [25]. These quantum numbers are described in more detail in section 3.3.7.

When the preliminary phosphine line list, SYT, was calculated, it was compared to the most recent experimental databases at the time. These were the CDMS [112] and HITRAN 2008 [140] databases, which have since been collated into the HITRAN 2012 database [139]. The HITRAN update has no impact on the discussion of results in section 3.4, since no experimental lines were modified or added, other than the inclusion of the CDMS data, which happened as a result of the publication of the article associated with this chapter. Throughout this chapter, the CDMS and the HITRAN databases will be referred to separately, as their analysis were made independently, and HITRAN will refer to HITRAN 08. Experimental comparisons to the final line list, SAlTY, described in Section 5, are made using the most current HITRAN database, HITRAN 2012.

The CDMS and HITRAN databases encapsulate decades of accurate laboratory measurements of phosphine spectra at room temperature. However, between the two databases, only a total of 22 230 lines are described, all within the $0-3600$ cm^{-1} region. CDMS contains transitions between $0 - 300$ cm^{-1} with J values of up to 34, sensitive to 10^{-36} cm/molecule, while HITRAN has transitions between $770 - 3600$ cm^{-1} stronger than 10^{-28} cm/molecule with $J \leq 23$. Only the rotational spectrum and a few bands have been covered and, most importantly, these databases are only designed to be valid for temperatures below 300 K. Other sources (see Table 3.1 below) mostly overlap with the HITRAN and CDMS databases, and add little more to these compilations. Recently, a paper by Malathy Devi *et al* improved the experimental phosphine spectrum for the $1950 - 2450$ cm^{-1} (5.13 – 4.08 μm) region [103], which is of particular importance due to it being an interval of low opacity in gas giants and brown dwarfs [123]. This was done by analyzing nine high-resolution, room-temperature absorption spectra recorded with two different Fourier transform spectrometers. These measurements were very accurately retrieved but nonetheless only consisted of a few thousand line positions and intensities, with no rotational excitations considered over J $= 14$. The data from Malathy Devi *et al* is included in the results comparison section of the SAlTY line list, in chapter 5.3.2. Collectively, all of these experimental databases for PH_3 contain less than 30 000 lines, and no wavelength coverage under 2.7 μm.

Table 3.1 presents an overview of the experimental spectra recorded for phosphine. This spans research starting from 1951 to the most recent body of work, published in 2014. The sources in the table are the result of an extensive, though possibly not exhaustive, literature review into the available phosphine data in the literature and experimental databases. The naming convention for each individual source consist of two digits for the date and first two letters of up to four authors' surnames, e.g. S. Yurchenko, M. Carvajal, P. Jensen, H. Lin, J. Zheng & W. Thiel 2005 becomes 05YuCaJeLi. The uncertainty column contains the stated uncertainty of the experiments, if available, and if the uncertainty varies between lines within the same article, "Yes" is written.

Table 3.1: Experimental sources of phosphine transition lines.

Source	**No of lines**	**Range(cm^{-1})**	**Uncertainties(cm^{-1})**	**Intensities**
13Muller[112][a]	2131	$0 - 292.8$	Yes	Yes
71DaNeWoKl[41][ac]	53	$0.0 - 4.8$	Yes	No
77HeGo[68][ac]	27	$0.0 - 17.8$	No	Calculated
81BeBuPoSh[17][ac]	68	$0.0 - 35.6$	No	No
51LoSt[96]	1	0.94	10^{-6}	No
81PiPoCo[132][c]	1	8.9	3.3×10^{-7}	No
69HeGo[67]	3	8.9 - 17.8	8.3×10^{-6}	No
06CaPu[31][a]	25	8.9 - 35.6	Yes	No
81BeBuGeKr[16][a]	52	10 - 35.7	No	No
74ChOk[35][ac]	12	$14.6 - 15.8$	2.7×10^{-5}	Calculated
79KrMeSk[77][c]	4	$17.1 - 17.8$	1.7×10^{-6}	No
00FuLo[54]	$155(+2500)$	$36 - 125(+750\text{-}1400)$	0.004	No
88FuCa[53][a]	118	$44.5 - 166.6$	0.002	No
02BrSaKlCo[24][b]	>1060	$770 - 1335$	0.0002	Some
81TaDaGo[158][b]	1318	$818 - 1340$	0.05	Calculated
97AiHaSpKr[2]	$62(+250^d)$	$924 - 1085$	6×10^{-6}	No
01HeZhHuLi[65]	28	$992 - 9040$	<0.01	Yes
04SaArBoWa[143]	26	$995 - 1093$	No	Yes
02UlBeKoZh[169]	≥ 4500	$1750 - 9200$	≤ 0.0005	No
92TaLaLeGu[159][b]	6605	$1885 - 2445$	0.0054	Yes
14DeKlSaBr[103]	1759	$1950 - 2450$	Yes	Yes
05WaChChDi[177]	1760	1950 - 2480 3280 - 3580	<6%	Yes
80BaMaNaTa[11]	1244	$2184 - 2446$	0.005	Approximate
02Suarez[155]	138	$2300 - 2381$	0.006	Yes
06BuSaKlBr[28][b]	8075	$2721 - 3601$	9.2×10^{-4}	Yes
07Kshiraga[78]	400	$2730 - 3100$	3.9×10^{-4}	No
73MaSaOl[102]	414	$2760 - 3050$	No	No
04UlBeKoZh[170]	>700	$3280 - 3580$	0.005	Transmittance

[a]Used in CDMS [112].

[b]Used in HITRAN 08 [140] and subsequent HITRAN 12 [139].

[c]Used in JPL[133].

[d]Compiled from other sources.

3.2.2 Previous Theoretical Work on Phosphine

There have been several previous attempts at *ab initio* studies of the behaviour and characteristics of phosphine spectra. Wang et al [176] computed a PES (potential energy surface) using coupled cluster CCSD(T) theory [121] and Dunning's cc-pVTZ basis sets (correlation consistent polarised Valence-only Triple Zeta) [46], followed by a vibrational analysis based on second-order perturbation theory. Wang et al determined several spectroscopic constants and fundamentals largely within $4\ \mathrm{cm}^{-1}$ of the observed results. CCSD(T) stands for Coupled Cluster theory [79] that explicitly treats Single and Double excitations but only treats Triple excitations perturbationally. This is the 'gold standard' method to obtain the PES and it is very good at describing the potential close to the equilibrium geometry, although increasingly unreliable towards dissociation. Zheng et al [65] calculated an *ab initio* three-dimensional P-H stretching DMS (dipole moment surface) using density functional theory and found absolute band intensities agreed with observed ones within a factor of 2. Yurchenko et al [186] presented a calculated, albeit empirically refined, PES and, using a variational method, calculated the rotational energy levels in the vibrational ground state of PH_3 for $J \leq 80$ [194], thus establishing for the first time the existence of six-fold near-degenerate ro-vibrational energy clusters for this molecule. Subsequently, Yurchenko et al [187] computed an entirely *ab initio* six-dimensional dipole moment surface (DMS) at the CCSD(T)/aug-cc-pVTZ (augmented cc-pVTZ using diffuse functions) level of theory for the electronic ground state of PH_3. This is the DMS that is used in the present work, and is described further in chapter 3.3.8 and in detail in [187]. The same paper also presented a PES determined by empirically refining an existing *ab initio* surface. Ovsyannikov *et al* [127] complemented the variationally calculated PES and DMS of Yurchenko *et al* [187] by calculating a PES at the CCSD(T) level using aug-cc-pV(Q+d)Z and aug-cc-pVQZ basis sets for P and H, respectively, and then presenting a list of computed vibrational transition moments for the electronic ground state of PH_3 [127]. In the Dunning basis set nomenclature Q stands for quadruple and d for the additional functions required by phosphorus. The PES used here is a refined version of the PES presented by Ovsyannikov *et al* [127], and is discussed further below. Nikitin *et al* produced a new PES [118] and used it to calculate several vibrational energy levels with only a marginal deviation from experiment. Nikitin *et al* then successfully modelled the lower three polyads of phosphine using a variational approach [117]. The resulting line list is discussed and compared to SYT in Section 3.4.5.

Despite these experimental and theoretical efforts, no complete line list for hot phosphine existed in the literature before the work described in this thesis was performed. The line lists mentioned above only cover low values of rotational excitations, a small number of bands, and some regions are missing altogether. It is not unusual for experimental data with extremely accurate line positions to have intensity uncertainties well above 10%, or no absolute intensities at all. Often, contamination of gas samples affects the measured intensities [103], and this can be an effect that is hard to accurately compensate for. Additionally, all the experimental datasets mentioned above are designed to be used only at room temperature and below. Consequently, they are insufficiently complete to be appropriate for use in the characterisation of atmospheres of any astronomical bodies at higher temperatures. The data presented in table 3.1 are nonetheless extremely useful, both as a tool to improve and refine the creation of a theoretical line list, but also as a way to validate its results.

3.3 Calculating the Cool Line List

3.3.1 TROVE

The production of the SYT phosphine line list presented here relies on the computational program TROVE (Theoretical ROVibrational Energies) [193], which uses a general variational approach to calculate the ro-vibrational energies for small semi-rigid polyatomic molecules of arbitrary structure. Solving the nuclear motion problem requires a choice of coordinates that is molecule specific, so TROVE achieves generality by including a coordinate transformation as part of its numerical procedure. TROVE computes the eigenfunctions for a specific molecule, as well as the associated ro-vibrational energies, and then calculates the line strengths for each of the transitions between energy levels. TROVE has been used to compute a large number of the ExoMol [165] molecules, including NH_3 [183], SO_3 [171], SbH_3 [188] and CH_4 [190]. For a more detailed description of the TROVE functionality, including its treatment of the kinetic and potential energy operators described below, the reader is directed to the TROVE paper [193].

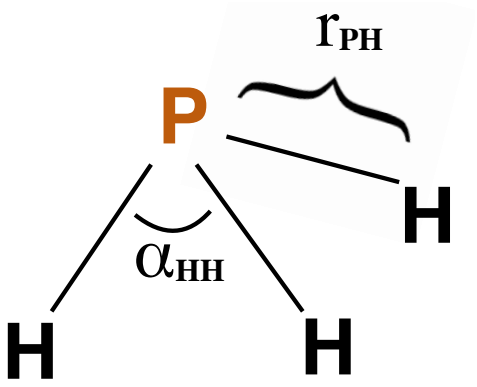
3.3.2 Kinetic and Potential Energy Operators

In TROVE, the nuclear kinetic energy operator is numerically constructed through a recursive scheme using a Taylor series expansion in terms of the molecule's internal coordinates, which requires no analytical pre-derivation, making its creation self-contained. This process is an extension of the XY3 program [189]. Although this procedure allows TROVE to simulate the nuclear motion for an arbitrary molecule, it does not calculate an exact kinetic energy operator. As such, the accuracy of the rotation-vibration energies depends on the expansion order of the approximate kinetic energy operator and its associated level of convergence. Similarly, the potential energy operator is expressed as a Taylor-type expansion, and its numerical derivation can lead to an accumulation of round-off errors. These approximate operators have been shown to converge as well as exact calculations for rigid, non linear molecules [134], while being considerably less computationally demanding.

3.3.3 Basis Functions

TROVE employs a numerical finite basis representation. This method employs six 1-D basis functions $\phi_i(\xi_i^\ell)$ $(i = 1 \ldots 6)$, where ξ_1^ℓ, ξ_2^ℓ, and ξ_3^ℓ are the linearized versions of the three stretching coordinates Δr_i, $i = 1, 2, 3$ $(\Delta r_i = \alpha_{\mathrm{PH}i} - \alpha_{\mathrm{e}})$, and ξ_4^ℓ, ξ_5^ℓ, and ξ_6^ℓ are the linearized versions of the three bending coordinates $\Delta\alpha_i$, $i = 1, 2, 3$ $(\Delta\alpha_i = r_{\mathrm{PH}i} - r_{\mathrm{e}})$, associated with the variation of the bond angles α_{23}, α_{13}, and α_{12}, respectively. Here r_{e} and α_{e} are, respectively, the equilibrium values for the bond length and the inter bond angle facing the bond length. Figure 3.3 illustrates the labelling of the bond lengths, r_{PH_i}, and the bond angles, $\alpha_{H_iH_j}$.

Figure 3.3: Labelling of the phosphine geometry.



The six basis functions, ϕ_i, can be separated into ϕ_{s_i} and ϕ_{b_i}, with $i = 1, 2, 3$, for the stretching modes and the bending modes, respectively.

To construct the ro-vibrational basis set, a contraction scheme described elsewhere [183, 184, 181] was followed. In order to control the size of the basis sets at all contraction steps the polyad-truncation scheme [193] is used, based on the PH_3 polyad number defined by

$$P = 2(s_1 + s_2 + s_3) + b_1 + b_2 + b_3 \leq P_{\max} \tag{3.1}$$

where s_i and b_i are the local quantum numbers associated with the basis functions, ϕ_{s_i} and ϕ_{b_i}. The basis set is formed only by the combination of basis functions ,ϕ_s and ϕ_b, for which $P \leq P_{\max}$.

The corresponding basis sets are formed as direct products of the primitive basis functions satisfying $2(s_1 + s_2 + s_3) \leq P_{max}$ (stretching) and $b_1 + b_2 + b_3 \leq P_{max}$ (bending) in agreement with Eq 3.1. so the maximal vibrational excitation is taken to be $P_{\max}/2$ and $P_{\max}$ for each of the stretching and bending modes, respectively.

The primitive basis functions for each vibrational mode ν_i, $i = 1 \ldots 6$, are generated by solving six 1-D Schrödinger equations associated with each mode using the Numerov-Cooley method [124, 37]. They are then used to solve two 3-D Schrödinger equations for each class of modes, stretching and bending, separately.

The resulting eigenfunctions, $\Phi_m^{\rm str}$ and $\Phi_n^{\rm bend}$, are then classified using C_{3v}(M) symmetry by analysing their transformational properties. The assigned local quantum numbers were translated into normal mode quantum numbers to allow for future ease of line assignment and comparison to the experimental spectroscopic data. The normal mode quantum numbers assigned to the eigenfunctions are $n_1, n_3^{L_3}$ (symmetric and asymmetric stretching modes) and $n_2, n_4^{L_4}$ (symmetric and asymmetric bending modes), where the $L_{3/4}$ quantum numbers account for the degeneracy of the asymmetric modes. The reassignment from local mode quantum numbers s_i,b_i $(i = 1 \ldots 3)$ to normal mode quantum numbers n_i $(i = 1 \ldots 4)$, $L_3 = |l_3|$ and $L_4 = |l_4|$ was performed by applying the following rules :

$$\begin{aligned} s_1 + s_2 + s_3 &= n_1 + n_3, \\ b_1 + b_2 + b_3 &= n_2 + n_4, \\ l_3 &= -n_3, -n_3 + 2, ..., n_3 - 2, n_3, \\ l_4 &= -n_4, -n_4 + 2, ..., n_4 - 2, n_4. \end{aligned}$$

It was also assumed that the symmetric modes ν_1 and ν_2 have lower energies than their asymmetric equivalents ν_3 and ν_4, respectively, and also that the vibrational energy grows when l_3 or l_4 increase. This set of rules was insufficiently sophisticated and prone to exceptions so the quantum number translation had to be done in a case-by-case basis. Fortunately, the reassignment of quantum numbers was performed at the initial $J = 0$ stage of the calculations where it could be extrapolated automatically, so only a small subset of the energies had to be translated by hand. At this stage, the reassignment is a reasonably straightforward process due to the relatively small number of solutions and simple polyad structure of the PH_3 energies: only approximately 700 sets of vibrational quantum numbers had to be translated at $J = 0$ for all energy levels ($J \leq 31$) to be assigned both sets of quantum numbers. For the production of the SAlTY line list, the TROVE calculations were updated to include both sets of quantum numbers automatically.

At the next contraction step, a 6-D vibrational basis is formed as a direct product of the stretching and bending wavefunctions $\Phi_m^{\rm str}$ and $\Phi_n^{\rm bend}$, contracted using the normal mode version of the polyad number truncation given by Eq 3.1:

$$P = 2(n_1 + n_3) + n_2 + n_4 \leq P_{\rm max}. \tag{3.2}$$

The 6-D functions $\Phi_m^{\rm str}$, $\Phi_n^{\rm bend}$ are then symmetrized [25], and used to solve the vibrational ($J = 0$) Schrödinger equation variationally by diagonalizing the ($J = 0$) Hamiltonian matrix on this 6-D vibrational basis and obeying the $P_{\rm max}$ polyad restriction from Eq 3.2. In this basis set, 1455, 1125, and 2571 basis functions were obtained for the A_1, A_2 and E symmetries, respectively.

At the last contraction step, the resulting eigenfunctions of the 6-D $J = 0$ problem are used to form our final rotation-vibrational basis functions in the so-called $J = 0$ representation [183] through a direct product with the symmetrized rigid rotor wave functions $|J, K, \tau\rangle$. In the $J = 0$ representation building the Hamiltonian matrix is straightforward as the vibrational part of the matrix is diagonal.

3.3.4 Convergence Tests

The accuracy with which high ro-vibrational states can be computed depends heavily on the size of the Hamiltonian matrix, which in turn depends on the size of the basis set.

To ensure that the size of the basis set was sufficiently large for the calculation of the energy levels, convergence tests were performed to facilitate the choice of the polyad number ($P_{\max}$), ranging from $P = 4$ to $P = 18$. A selection of the results from the convergence tests is given in table 3.2.

Table 3.2: $J = 0$ energy eigenvalue convergence with polyad number, $P_{\max}$, used to generate the basis set, where $P(n) = (P_{max} = 18) - (P_{max} = n)$.$N/A$ stands for not available in the literature.

Band	**Experiment**	**$P_{\max}$ =18**	**P(16)**	**P(14)**	**P(12)**	**P(10)**	**P(8)**
ν_0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ν_2	992.135	992.732	-0.010	-0.033	-0.090	-0.243	-0.714
$2\nu_2$	1972.571	1967.427	-0.133	-0.393	-0.939	-2.197	-5.321
$2\nu_4$	2226.835	2222.257	-0.009	-0.033	-0.095	-0.283	-0.897
ν_1	2321.121	2324.077	-0.003	-0.011	-0.036	-0.128	-0.436
$3\nu_2$	2940.858	2911.866	-1.325	-3.472	-7.230	-14.335	-29.377
$\nu_2 + 2\nu_4$	3214.936	3204.920	-0.115	-0.361	-0.941	-2.467	-7.220
$\nu_1 + \nu_2$	N/A	3322.008	-0.038	-0.132	-0.384	-1.135	-3.820
$3\nu_4$	N/A	3341.959	-0.028	-0.098	-0.295	-0.931	-3.434
$\nu_3 + \nu_4$	3440.258	3439.304	-0.009	-0.034	-0.117	-0.437	-1.843
$4\nu_2$	N/A	3803.487	-9.750	-22.284	-39.911	-66.534	-108.521
$2\nu_2 + 2\nu_4$	N/A	4159.298	-1.025	-2.839	-6.309	-13.381	-28.564
$\nu_1 + 2\nu_2$	N/A	4303.413	-0.352	-1.074	-2.740	-6.742	-16.622
$\nu_2 + 3\nu_4$	N/A	4318.966	-0.257	-0.786	-1.981	-4.906	-12.643
$\nu_2 + \nu_3 + \nu_4$	N/A	4410.959	-0.131	-0.436	-1.196	-3.221	-9.642
$4\nu_4$	N/A	4422.841	-0.098	-0.325	-0.919	-2.707	-7.461

One would usually expect the chosen polyad number $P_{\max}$ to be the one which provides full convergence; as can be seen from table 3.2, this is not achieved even for the highest polyad configuration, $P_{\max} = 18$. However, although the calculations should improve as the polyad number increases, since the 'spectroscopic' PES (see Section 3.3.5) used to obtain the energy levels was generated with $P_{\max} = 14$, the values for the calculated energy levels move away from the observed values when a higher polyad configuration is used. The PES used in this work relies on a self consistency of parameters, and as such it is only an effective potential for a specific set of parameters, so $P_{\max} = 14$ was chosen.

Table 3.2 was meant to test the convergence of energy values only as a function of polyad number, so it was done using expansions of fourth order for both the kinetic energy operator and the potential energy operator; this accounts for the significant discrepancies between the computed band origins and the corresponding experimental values. The expansion orders for both operators were significantly increased for the final SYT calculations, and even further for SAlTY.

Multiple convergence tests were performed to test the expansions of the kinetic and potential energy operator and, although accuracy of the energy level values did improve considerably as the expansions grew, so did the computational cost involved. Consequently, the expansions of the kinetic energy operator and potential energy function were, respectively, truncated at the 6^{th} and 8^{th} order, which provide a good convergence level and were, at the time, the highest possible orders that did not make the project computationally prohibitive, particularly for high rotational excitations.

3.3.5 Potential Energy Surface

The PES of phosphine used here is a refinement of the aforementioned *ab initio* (CCSD(T)/aug-cc-pV(Q+d)Z) PES [127], done by performing a least square fit to available experimental ro-vibrational energies with $J = 0, 1, 2, 4$ and 10. The refinement followed the fitting procedure introduced in Ref. [184], where the modification to the PES is represented as a correction ΔV to the *ab initio* PES, V_0. In this refinement, the ro-vibrational $J = 0, 1, 2, 4, 10$ eigenfunctions of the 'unperturbed' Hamiltonian $H_0 = T + V_0$ are used as basis functions when solving the set of Schödinger equations for the ro-vibrational behaviour of the molecule. This process is done iteratively with a variable ΔV in order to minimise the

associated functional in the least-squares fitting. Then an extended Hamiltonian, H', can be represented as $H_0 + \Delta V$ where ΔV is the off-diagonal part of the Hamiltonian.

With the refined PES, a total root-mean-squares (rms) error of 0.03 cm^{-1} was obtained for the fitting energy set. Table 3.3 shows the change in the band centre values between the pre- and post-refinement PES.

	Band Centres (cm^{-1})		
Band	**Obs**	**Calc '06 [187]**	**Present**
ν_0	0.000	0.000	0.000
ν_2	992.130	992.500	992.152
ν_4	1118.310	1117.870	1118.322
$2\nu_2$	1972.550	1972.820	1972.590
$\nu_2 + \nu_4$	2108.150	2107.170	2108.169
$2\nu_4$	2226.830	2227.860	2226.835
$2\nu_4$	2234.930	2234.570	2234.940
ν_1	2321.120	2322.040	2321.142
ν_3	2326.870	2329.180	2326.888

Table 3.3: Observed and calculated band centres, with "Calc '06" referring to calculations done with the original potential [187] and "Present" to the SYT calculations done using the refined PES.

As already mentioned in sections 3.3.2 and 3.3.4, the kinetic energy operator is not calculated exactly and the basis set is artificially truncated. Thus the refined PES must be considered as an effective PES; it only gives the accurate results presented in Table 3.5 when used in conjunction with TROVE and the parameters described above (operator expansion order and polyad number). The refined PES used for SYT is available online at *www.exomol.com* and as supplementary material to the associated article.

The PES is a major source of error in the line list creation process, and as such constant updates and refinements are necessary. Although the PES used in SYT is appropriate for creating an accurate room temperature line list with wavenumbers up to 8000 cm^{-1} and $J \leq 31$, it would not necessarily remain so for a line list with parameters exceeding this. In fact, the creation of the final phosphine line list required a further refinement to the PES, to accommodate a different set of parameters necessary for the creation of a hot line list.

The PES used in SYT also fails to accurately describe the potential of phosphine at very high excitations, showing some artificial minima in the region of dissociation. Consequently, it is not suitable for use in dynamical simulations without adjustment.

3.3.6 Band Centre Replacement

Although using an experimentally refined, or 'spectroscopic', PES improves the values of the energy levels, the TROVE calculations still do not completely reproduce the observed transition frequencies. To correct this, a further empirical approach was adopted, where an artificial frequency shift is added to calculated band origins as given by the $J = 0$ energies; this procedure is described as an empirical basis set correction scheme (EBSC) [185]. It leads to a rotational energy structure in much better agreement with experimental results not just for the $J = 0$ energies, but for the remaining J values of the band as well. As a test, the $J = 1, 2$ and 4 values were compared to those from experiments which, together with $J = 0$ were also the values of J used to refine the PES.

While using EBSC, care must be taken to only select reliable experimental data, since data of limited accuracy can be a source of error which TROVE cannot compensate for. Small shifts to the band origin can have repercussions not just throughout their band but also in neighbouring bands, so an exact replacement of the theoretical band origins with their experimental equivalent did not result in an overall improvement. Instead dozens of iterative shifts to the band origins were performed, in an attempt to maximise the improvement of the energy values within all the bands for which there were reliable experimental data for comparison.

Once no significant overall improvement could be gained through the EBSC, a total of 11 band centres were replaced with the values used in the most optimized iteration. The final improvements to the standard deviation, σ, of the whole band from the experiment values are displayed on table 3.4. The original average σ value for the set of un-shifted bands was 0.031 cm^{-1} which was reduced to 0.020 cm^{-1} after the band centre replacement.

Band	Obs	Original σ	New σ
ν_2	992.135	0.030	0.020
ν_4	1118.307	0.012	0.005
$2\nu_2$	1972.571	0.011	0.007
$\nu_2 + \nu_4$	2108.152	0.045	0.034
$2\nu_4$	2226.835	0.029	0.010
$2\nu_4$	2234.920	0.032	0.014
ν_1	2321.121	0.028	0.012
ν_3	2326.8667	0.023	0.013
$3\nu_2$	2940.767	0.063	0.046
$\nu_2 + 2\nu_4$	3214.936	0.035	0.024
$\nu_3 + \nu_4$	3440.259	0.037	0.034

Table 3.4: Observed band origins (Obs) and standard deviation with which TROVE reproduces the terms within each band, σ, before and after replacement of the band origins. Observed data is from HITRAN2008 [140]. Values in the "Original σ" column obtained with the refined, not the *ab initio*, PES.

3.3.7 Energy State Quantum Numbers

The chosen ranges of energy eigenvalues for the room temperature linelist were 4000 cm^{-1} for the highest lower energy and 12000 cm^{-1} for the highest upper energy. These choices allow for a complete set of transitions in the range 0 - 8000 cm^{-1} for temperatures up to 300 K. With these parameters, 5.6 million rotation-vibration energy levels were computed.

To fully describe an energy level corresponding to an internal rovibrational motion of phosphine one needs the following minimal set of quantum numbers [84]:

$$n_1, n_2, n_3, n_4, L_3, L_4, L, \Gamma_{\rm vib}, J, K, \Gamma_{\rm rot}, \Gamma_{\rm tot}, \tag{3.3}$$

- The vibrational quantum numbers n_1 (symmetric stretch), n_2 (symmetric bend), n_3 (asymmetric stretch) and n_4 (asymmetric bend) correspond to excitations of, respectively, the ν_1, ν_2, ν_3 and ν_4 vibrational modes. The doubly degenerate modes ν_3 and ν_4 require additional quantum numbers $L_3 = |l_3|$ and $L_4 = |l_4|$, which describe the projections of the corresponding angular momenta (see, for example, Ref. [25]).

- The projection of the total vibrational angular momentum, $L = |l|$, characterizes the coupling of l_3 and l_4 and was only added to the quantum number assignment at the end of the calculations, in an attempt to further reduce ambiguity in the description of the energy levels. This quantum number, L, was added with the following rules: L must be a multiple of 3 if $\Gamma_{\rm vib} = A_1, A_2$ and cannot be a multiple of 3 if $\Gamma_{\rm vib} = E$. By convention, the lower energy value is assigned to $|L_3 - L_4|$, and the highest to $|L_3 + L_4|$. However, at higher energies, TROVE does not necessarily assign unique quantum labels to every state. In this case all states are assigned to $|L_3 - L_4|$.

- $\Gamma_{\rm rot}$, $\Gamma_{\rm vib}$, and $\Gamma_{\rm tot}$ are, respectively, the symmetry species of the rotational, vibrational and total internal wavefunctions in the molecular symmetry group $C_{3\rm v}$(M), spanning A_1, A_2 and E. The spatial symmetry label is given by a molecule's point group symmetry rather than by the projection onto the nuclear axis.

- Finally, J is the total angular momentum and $K = |k|, k = -J, \ldots, J$ is the projection of the total angular momentum on the molecule fixed axis z.

Selection Rules

Out of the twelve quantum numbers in 5.1, only two are rigorously conserved quantum numbers (J and $\Gamma_{\rm tot}$). This quantum number description is similar to that proposed by Down *et al* [84] for ammonia.

The rigorous dipole selection rules allow transitions between states with $\Gamma_{\rm tot}$ satisfying $A_1 \leftrightarrow A_2$ and $E \leftrightarrow E$ and $\Delta J = 0, \pm 1$ only. Strong transitions also obey the dipole selection rules $\Delta J = \pm 1$ and $\Delta K = 0$, but there are weakly allowed transitions which occur when, due to centrifugal distortion, the C_{3v} geometrical symmetry of the molecule is broken and K is no longer a good quantum number [35]. Only approximate selection rules can be associated with the normal mode vibrational quantum numbers n_1, n_2, n_3, n_4.

3.3.8 Dipole Moment Surface

An existing six-dimensional *ab initio* electric dipole moment surface (CCSD(T)/aug-cc-pVTZ) [187] was used to obtain the Einstein coefficients and transition intensities for both the preliminary line list SYT and SAlTY. The DMS was calculated on a large grid of 10080 molecular geometries. The procedure used to compute absolute intensities mirrored that used

for ammonia, and is described in detail in Yurchenko *et al* [185]. A number of vibrational transitional moments were calculated to help characterize the quality of the DMS. As mentioned in 2.3.2, transition moments are also affected by the accuracy of the wavefunctions and hence the accuracy of the underlying PES.

Table 3.5 compares empirical vibrational transition moments for phosphine with values computed by TROVE, both in the present work and in the previous studies [187]. It is clear that the theoretical band intensities are in good agreement with experiment and, although the DMS used here is unchanged from that used in Yurchenko *et al* [187], the use of our refined PES and consequently improved wavefunctions has meant that our new results have generally reduced the error in the calculations of the transition moments, from an average deviation of 22.5% to 10%.

	Band Centres (cm^{-1})		Transition Moments (D^2)		
Band	**Obs**	**Present**	**Obs**	**Calc '06 (%) [187]**	**Present (%)**
ν_0	0.000	0.000	0.57395(30)	0.583(1.6%)	0.585(2.0%)
ν_2	992.130	992.152	0.08251(5)	0.085(2.5%)	0.084(2.2%)
ν_4	1118.310	1118.322	0.08626(5)	0.087(0.3%)	0.089(3.7%)
$2\nu_2$	1972.550	1972.590	0.00299(5)	0.003(9.7%)	0.004(21.4%)
$\nu_2+\nu_4$	2108.150	2108.169	0.01102(6)	0.009(15.6%)	0.014(25.3%)
$2\nu_4$	2226.830	2226.835	0.0176(2)	0.006(68.8%)	0.018(1.5%)
$2\nu_4$	2234.930	2234.940	0.0176(2)	0.002(91.5%)	0.013(23.5%)
ν_1	2321.120	2321.142	0.0690	0.073(6.1%)	0.072(4.3%)
ν_3	2326.870	2326.888	0.130	0.139(6.8%)	0.138(6.1%)

Table 3.5: Observed and calculated band centres and their respective transition moments (deviation from experimental values shown as a percentage). Uncertainties of the experimental (Obs) transition moments are given in parentheses (in units of the last digit quoted) where available.

3.4 Cool Phosphine Results

At the end of the SYT calculations, a room temperature ro-vibrational spectrum for phosphine was produced, covering the wavenumber range 0 to 8000 cm^{-1}. The associated line list consists of a total of 137 million transitions between 5.6 million energy levels.

One of the most reliable ways of validating a theoretical line list is to compare it to existing high accuracy experimental measurements. Once the SYT line list was completed, its transition frequencies were compared to that of databases of experimentally measured lines. To this end, an extensive survey of the literature was performed to locate all the experimental spectra recorded for phosphine, the results from which can be seen in Table 3.1 in Section 3.2.

Multiple computational algorithms were attempted in an effort to automate the comparison between theoretical lines and those from laboratory sources. These mostly relied on comparing lines with similar characteristics, namely quantum numbers. However, due to the ambiguity of experimental assignment of quantum numbers, together with the insufficient description of many transition lines, this was not possible. Instead, a line-by-line combination differences analysis was performed for both the CDMS [112] and the HITRAN [140] databases. Combination differences is a process in which each upper and lower experimental energy level is matched with its theoretical equivalent by comparing the quantum numbers of the states and their energy frequency. Then experimental transitions can be compared to the calculated transitions by subtracting between matched energy pairs. However, due to the inconsistency and incomplete labelling of the available experimental data, this was impossible to do in an automated fashion. Consequently, thousands of the experimental energies were matched by hand, and only the final transition match step was able to be automated, using MATLAB [62]. This process of combination differences is only possible because there are relatively few energy levels in the experimental databases, but it nonetheless required months of continuous work.

The combination differences analysis allowed for a reliable comparison between the calculated data and accurate experimental measurements. The SYT line list agreed very well the observed phosphine spectra at room temperature, with a maximum rms (root mean square) deviation from CDMS of 0.076 cm^{-1} for the rotational spectrum and of 0.23 cm^{-1} from HITRAN. Problems with ambiguous or incorrect labelling of energy levels in both databases skewed the rms results and their true value is expected to be significantly lower. To test this,

the worst 1% matches were removed from the comparison; as predicted, the rms deviations lowered significantly to 0.05 cm^{-1} for CDMS and 0.19 cm^{-1} for HITRAN.

Additional causes for the possible artificial inflation of the rms value are the experimental assignment of the same quantum numbers to experimental states with different energies, and energy level values obtained with only one or few more measurements, making them highly unreliable for comparison.

3.4.1 Energy Eigenvalues

For the creation of the SYT line list the following thresholds were selected.

The chosen ranges of energy eigenvalues are 4000 cm^{-1} for the highest lower energy and 12000 cm^{-1} for the highest upper energy. These choices allow for a complete set of transitions in the range 0 - 8000 cm^{-1} for temperatures up to 300 K. A completeness analysis of the phosphine line list with respect to temperature is discussed further in chapter 5.3.

Performing an explicit summation over all the energy levels calculated for SYT yields a partition function equal to 3249.5 at 296 K. This is 0.028% higher than the latest published HITRAN value of 3248.6 [152]. Further work on the partition function and other thermodynamic properties of both phosphine and ammonia can be found in the following chapter, 4.

The lower energy threshold of 4000 cm^{-1} defines the highest J that had to be taken into account, as $J = 31$ is the highest value of J for which eigenvalues below 4000 cm^{-1} exist. In fact, at $J = 31$, only a total of 8 levels, across all symmetries, had energies below 4000 cm^{-1}, corresponding to 0.0025% of all energy eigenvalues for that J value. Figure 3.4 shows the variation in number of useful energies (those below 4000 cm^{-1}) as J increases.

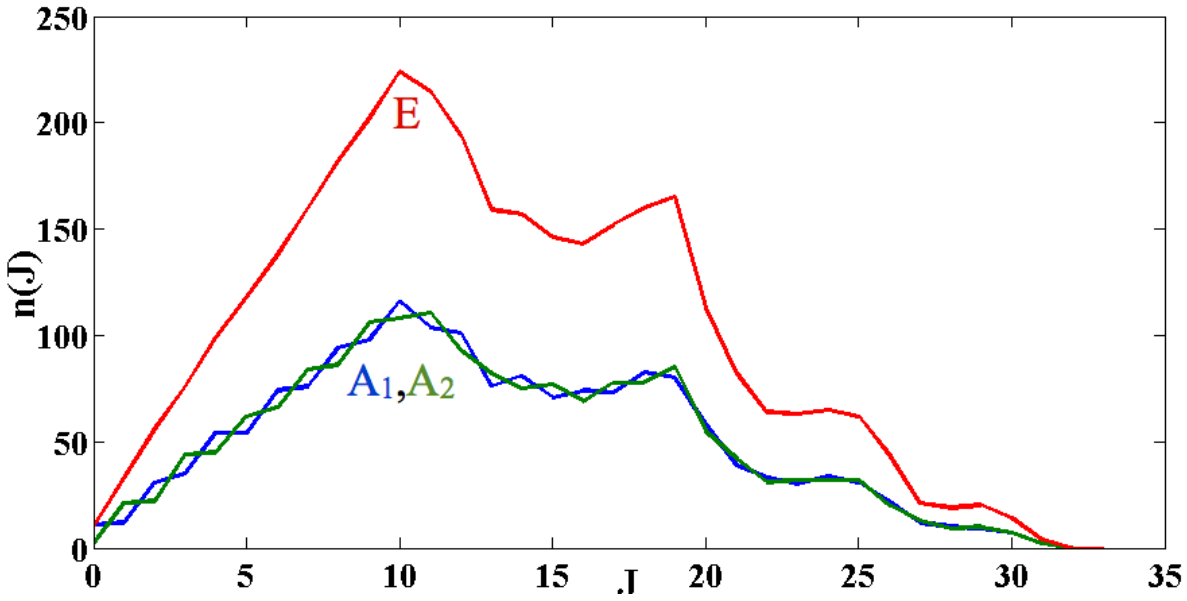


Figure 3.4: Total number of energy eigenvalues under 4000 cm^{-1} for $J = 0, 1, ..., 33$. The blue, green and red lines represent, respectively A_1, A_2 and E symmetry eigenvalues.

3.4.2 Overview

Phosphine has a clear spectrum that does not get sharply weaker with growing wavenumber at room temperatures. The vibrational bands that most contribute to its shape are described in table 3.6.

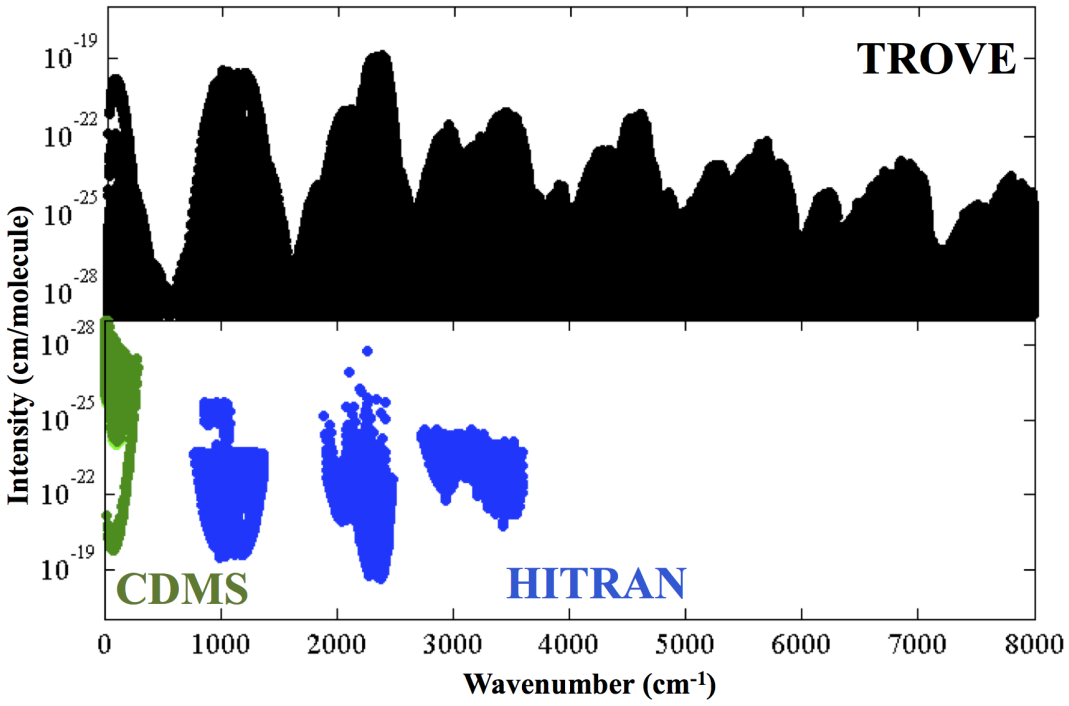
For the purposes of comparison with the existing experimental data, only transitions with an intensity above a minimum threshold were included. In the region $0 - 300$ cm^{-1}, only transitions stronger than $\geq 10^{-13}$ cm/mol (or 1.7×10^{-37} cm/molecule) were considered, to compare with the CDMS [112] database. For transitions with wavenumbers ≥ 300 cm^{-1} this threshold was reduced to 10^{-5} cm/mol (1.7×10^{-29} cm/molecule) to compare to the HITRAN 2008 [140] database. These thresholds are approximately one order of magnitude more sensitive than the weakest lines in both experimental databases. With this intensity cut-off, only a selection of 5 488 177 transitions were used in the comparison to experiment, from a total of 137 255 400 computed lines.

Figure 3.5 shows an overview of our simulation compared to the CDMS [112] and HITRAN [140] databases. HITRAN lacked any pure rotational transitions until the 2012 addition of the CDMS rotational spectrum at our suggestion [139]. Additionally, although HITRAN (08 and 2012) captures most lines with intensities at room temperature stronger than 5×10^{-25} cm/molecule, it is very incomplete for weaker lines.

Table 3.6: Dominant bands in the room temperature spectrum of phosphine.

Mode	Band Centers (cm^{-1})	Description
$2\nu_2 - \nu_2$	$1972.550 - 992.130 = 980.42$	'Hot' band
ν_2	992.130	Fundamental symmetric bending
ν_4	1118.310	Fundamental asymmetric bending
$2\nu_2$	1972.550	Fundamental overtone
$\nu_2 + \nu_4$	2108.150	Combination band
$2\nu_4$	2226.830	Fundamental overtone
ν_1	2321.120	Fundamental symmetric stretching
ν_3	2326.870	Fundamental asymmetric stretching
$4\nu_2 - \nu_2$	$3895.685 - 992.130 = 2903.555$	'Hot' band
$3\nu_2$	2940.810	Fundamental overtone
$2\nu_2 + \nu_4$	3085.609	Combination band
$\nu_1 + \nu_2$	3307.624	Combination band
$\nu_2 + \nu_3$	3311.959	Combination band
$\nu_1 + \nu_4$	3424.627	Combination band
$\nu_3 + \nu_4$	N/A	Combination band
$\nu_2 + 2\nu_4$	N/A	Combination band

Figure 3.5: Comparison between the simulated absorption spectrum created by TROVE and that from the CDMS[112] and HITRAN [140] databases, as a function of wavenumber.



All the information on the 5.6 million rotation-vibration energy levels and the associated list of the 137 million transitions was made available in the published paper, as well as the ExoMol website *www.exomol.com*, in the ExoMol format [164]. The files for the energy levels, transitions and HITRAN format line list have now been replaced by the upgraded phosphine line list, SAlTY, described in chapter 5. The refined PES and DMS used for SYT is provided as a Fortran program in the supplementary to the article associated with this chapter.

3.4.3 Rotational Spectrum

Experimental data for the rotational spectrum was obtained from the CDMS database [112], although the strongest of these lines can also be found in the JPL database [133, 130]. The documentation for the CDMS data can be found online and cites its sources as Cazzoli and Puzzarini [31], Belov *et al* [16], Fusina and Carlotti [53], Davis, Newman, Wofsy, and Klemperer [41], Chu and Oka [35], Helms and Gordy, [68], and Belov, Burenin, Polyanski, and Shapin [17]. Further information about these papers can be found in Table 3.1. The CDMS data set contains 2131 transition lines in the region 0 - 300 cm^{-1} with a maximum J quantum number of 34.

A combination differences analysis was done to adequately compare the lines calculated for SYT to those in CDMS; the lower and upper energy levels were matched between the datasets and the difference between the resulting transitions analysed. Due to the labelling ambiguity (between A_1 and A_2) in many of the energy levels with A symmetries, only the E symmetry transitions were matched with sufficient confidence for an accurate analysis. The E-symmetry sample can be expected to give a representative rms for the remaining transitions. Combination differences, as described above, together with an algorithm created with MATLAB, matched theoretical lines produced by TROVE with equivalent ones in the CDMS dataset. These matches deviated from those in CDMS with a rms value of 0.076 cm^{-1}. As mentioned earlier, this value is expected to be overestimated.

Figure 3.6 shows close-ups of the rotational spectrum, where a line by line comparison can be made with the CDMS data mirroring TROVE's theoretical predictions, and those transitions that could be matched with confidence highlighted in red.

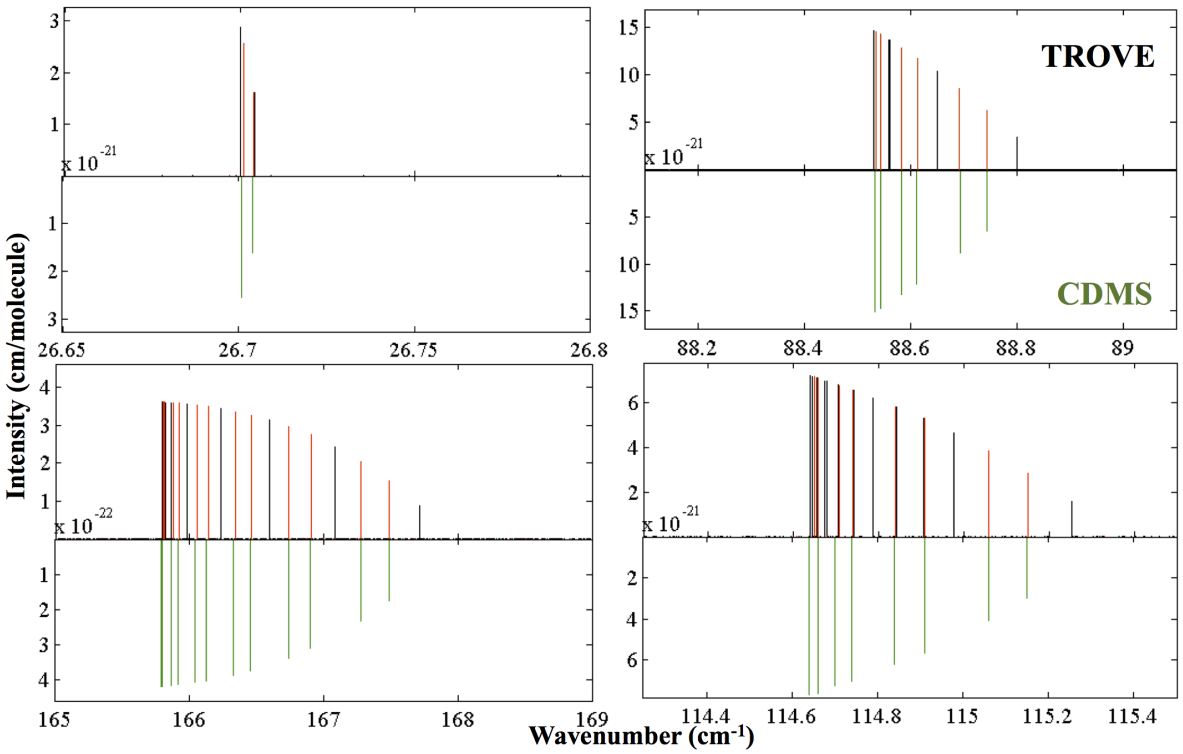


Figure 3.6: Rotational Spectrum showing CDMS (lower) versus TROVE (upper). The matched TROVE transitions are highlighted in red.

3.4.4 HITRAN Database

The 2008 HITRAN data [140] that was used to compare to the room temperature for phosphine contains 20 099 transitions in the region 770 - 3601 cm^{-1}. Of these, only 16 008 are assigned and of these 2011 have no upper vibrational quantum numbers, possibly due to vibrational mixing. There seems to be no consistent description of symmetry and many transitions appear either mislabelled or misassigned. Line mixing makes measurements and assignments difficult, and is particularly problematic for A symmetry transition pairs with $\Delta K = 3n$, $n = 1, 2, 3....$

The energy levels of the assigned lines were compared to those calculated in TROVE and this information was used to match lines between experiment and theory. Similarly to the CDMS dataset, ambiguity between energy levels with A symmetry meant that only E symmetry transitions were considered in the comparison. Again, this sample can be expected to give a representative rms for each band. Assigned E symmetry transitions correspond to 7838 of the total HITRAN lines. Thirteen of these transitions had ambiguous or incorrect K values, so only 7825 E symmetry transitions were matched. These were found to have a

rms deviation of 0.23 cm^{-1} from those calculated by TROVE. The wavenumbers of a few transitions involving energy levels which are defined by a single transition in the HITRAN database disagree significantly with that of those calculated with TROVE. These skew the rms deviation and again, its true value is expected to be significantly lower. When the worst 1% matches are removed from the comparison, the rms deviation lowers to 0.19 cm^{-1}. Values for the rms deviation for individual bands are found in the following sections.

The regions of the non rotational section of the spectrum can be divided into polyad number, as per eq 3.2. The bending modes ν_2 and ν_4 each contribute 1, while the stretching modes ν_1 and ν_3 each contribute 2 (e.g. P($\nu_3 + \nu_4$)= P($4\nu_2 - \nu_2$) = 3).

P = 1

The first region (Polyad number = 1) is located between 770 – 1372 cm^{-1} and is dominated by the fundamental bending bands ν_2 and ν_4 and the 'hot' band $2\nu_2 - \nu_2$. TROVE's rms deviation from HITRAN in this region is 0.23 cm^{-1}. Its rms is 0.37 cm^{-1} for the ν_2 band, 0.11 cm^{-1} for the ν_4 band and 0.11 cm^{-1} for the $2\nu_2 - \nu_2$ band. HITRAN references Brown, Sams and Kleiner [24] as the source for this region. Figure 3.7 shows close-ups of a selection of representative sub-regions within the $P = 1$ region.

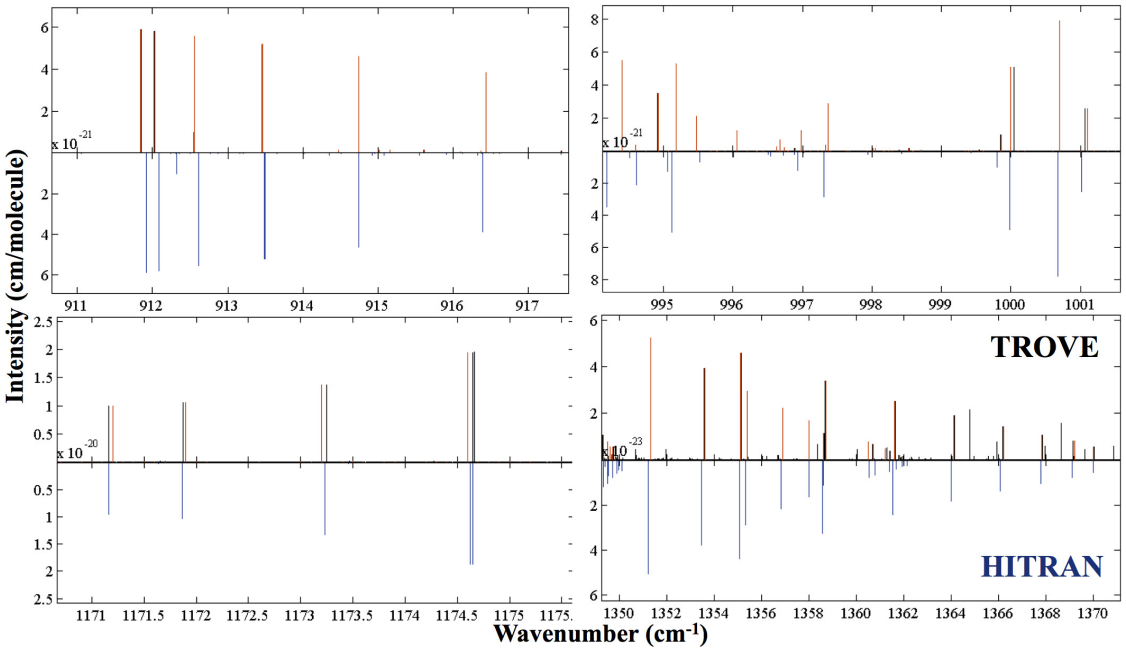


Figure 3.7: HITRAN (lower) versus TROVE (upper), with matched transitions highlighted in red, for the region $P = 1$.

P=2

The pentad region (45 μm or 1950 - 2450 cm^{-1}) is dominated by a hot $\nu_2 + \nu_4$ weak band, and two dominant and overlapping bands: the $2\nu_4$ overtone and fundamentals ν_1 and ν_3. The two fundamentals experience a strong Coriolis interaction.

The second region is located between 1885 – 2478 cm^{-1} and is dominated by the fundamental stretching bands ν_1 and ν_3, the $2\nu_2$ and $2\nu_4$ fundamental overtones and the combination band $\nu_2 + \nu_4$. TROVE's rms deviation from HITRAN in this region is 0.20 cm^{-1} overall, with 0.28 cm^{-1} for the ν_1 band, 0.22 cm^{-1} for ν_3, 0.05 cm^{-1} for $2\nu_2$, 0.15 cm^{-1} for $2\nu_4$ and 0.08 cm^{-1} for $\nu_2 + \nu_4$. HITRAN references Tarrago *et al* [159] as the source for this region. Figure 3.8 shows close-ups of a selection of representative sub-regions within the $P = 2$ region.

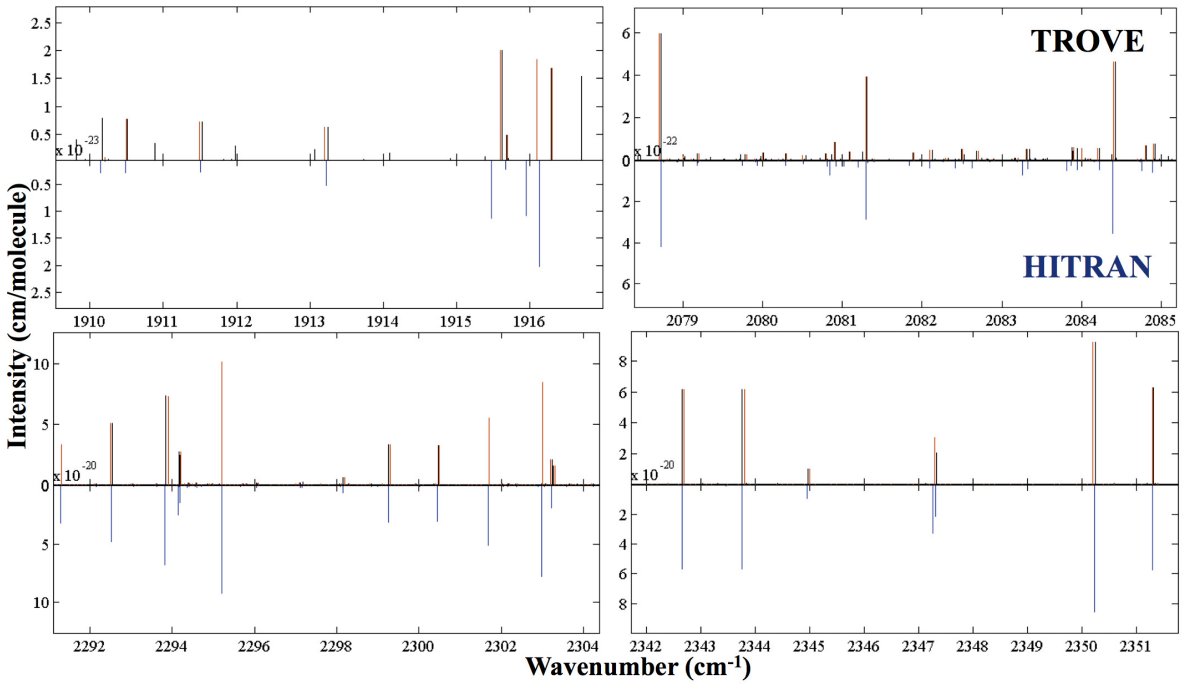


Figure 3.8: HITRAN (lower) versus TROVE (upper), with matched transitions highlighted in red, for the region $P = 2$.

P=3

The third region is located between 2733 – 3600 cm^{-1} and is dominated by the $4\nu_2 - \nu_2$ hot band, the $3\nu_2$ overtone and the $\nu_1 + \nu_2$, $\nu_1 + \nu_4$, $\nu_2 + \nu_3$, $\nu_3 + \nu_4$, $\nu_2 + 2\nu_4$ and $2\nu_2 + \nu_4$ combination bands. TROVE's rms deviation from HITRAN in this region is 0.27 cm^{-1}, with 0.46 cm^{-1} for $4\nu_2 - \nu_2$, 0.37 cm^{-1} for $3\nu_2$, 0.50 cm^{-1} for $\nu_1 + \nu_2$, 0.21 cm^{-1} for $\nu_1 + \nu_4$, 0.23 cm^{-1} for $\nu_2 + \nu_3$, 0.19 cm^{-1} for $\nu_3 + \nu_4$, 0.11 cm^{-1} for $\nu_2 + 2\nu_4$ and 0.28 cm^{-1} for $2\nu_2 + \nu_4$. HITRAN references Butler *et al* [28] as the source for this region. Figure 3.9 shows close-ups of a selection of representative sub-regions within the $P = 3$ region.

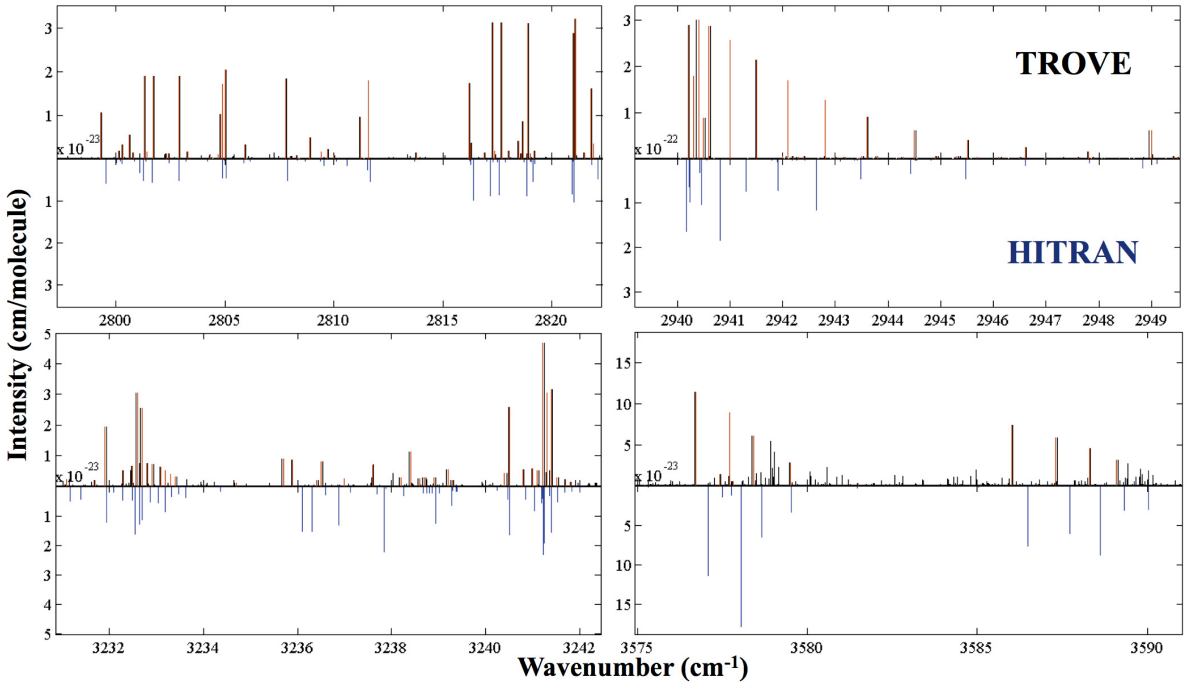


Figure 3.9: HITRAN (lower) versus TROVE (upper), with matched transitions highlighted in red, for the region $P = 3$.

3.4.5 Other Comparisons

Nikitin *et al* [118], also using a theoretical variational approach, computed 55 223 transitions in the range 700 – 3500 cm^{-1}, which is similar to the HITRAN 08 range [140]. Their line list does not aim for completeness and even within the HITRAN regions appears to have significant omissions (e.g. maximum $J = 20$). SYT improves on their root mean squares deviation from observed band centres of 1.4 cm^{-1}.

Experimental data (1768 lines) from Wang *et al* [177] in the 1950-2480 cm^{-1} and 3280-3580 cm^{-1} regions was compared to the theoretical lines created by TROVE, and the two are in good agreement. The results presented here deviated from those of Wang *et al*'s with an overall rms deviation of 0.17 cm^{-1}, for those lines that could be matched with confidence. The rms deviation for the first region was 0.11 cm^{-1}, compared to 0.20 cm^{-1} for the equivalent HITRAN region and 0.21 cm^{-1} for the second, compared to 0.27 cm^{-1} for HITRAN.

Additionally, our intensities agree significantly better with the Wang data than with HITRAN's data, as can be seen by the intensity plots in Figure 3.10, suggesting that Wang *et al*'s intensities should be used in a future release of HITRAN. Recently, the region of polyad number 2 was revisited by Devi *et al* [103]. This did not add any further lines to the region, but improved the intensities of the transitions. These were not compared to the SYT line list, but comparisons with SAlTY (see chapter 5.3.2) show a marked improvement in that region.

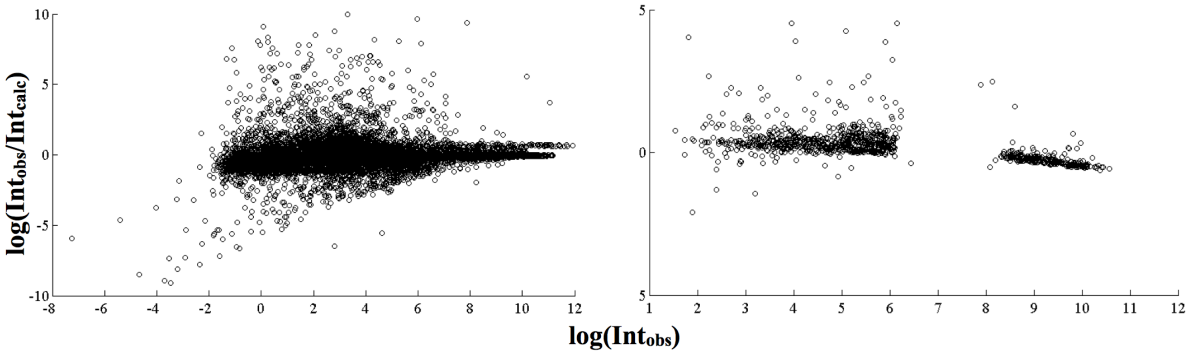


Figure 3.10: Plot of $log(I_{obs}/I_{calc})$ versus $log(I_{obs})$ for the HITRAN data [140] (left) and the Wang data [177].

The JPL data [130, 133] also agrees very well with SYT. It contains 729 transitions between $0 - 188$ cm^{-1} with J$\leq$ 27. Combination differences analysis of the E symmetry transitions showed that the data presented here deviated from JPL with a rms of 0.066 cm^{-1}.

When the worst 1% matches are removed from the comparison, the rms deviation lowers to 0.031 cm^{-1}. Overall, the JPL transition wavenumbers agree more closely with those presented here than CDMS' (0.076 cm^{-1} and 0.05 cm^{-1} with and without the worst 1% matches, resepctively). However, the hundred strongest transitions in the rotational spectrum from CDMS match ours much better (0.003 cm^{-1} compared to JPL's 0.07 cm^{-1}). Additionally, CDMS is considerably more complete and its intensities are in closer agreement with ours. This should be, at least partially, due to the fact that some of the K=3n transition doublets are very close so their intensities are combined in the JPL database, while the equivalent TROVE lines are left resolved. This difference in data handling leads to an apparent intensity disagreement by a factor of two.

3.5 Discussion

The SYT line list contains 137 million transitions between 5.6 million energy levels for rovibrational states up to $J_{max} = 31$ and wavenumber range 0-8000 cm^{-1}. This work replicates very well the observed phosphine spectra at room temperature, with a maximum rms deviation from CDMS of 0.076 cm^{-1} for the rotational spectrum and of 0.23 cm^{-1} from HITRAN. As a result of the published paper associated with this chapter, the CDMS database was incorporated in what is now the most current HITRAN database, HITRAN 2012 [139].

SYT is a valid line list for any phosphine analysis below 300 K; however the hot line list described in chapter 5, SAlTY, is capable of simulating observed spectra from astronomical bodies at higher temperatures ($\leq$ 1500 K) and should be used for applications at all temperatures, as it is substantially more complete.

The creation of the SYT line list tested the limits of the available computational facilities. The production of the final, hot line list for phosphine required even more computational power and memory, as it necessitates bigger basis sets and denser eigenvectors. The size of the Hamiltonian matrices requiring diagonalization grows as (2J + 1); the difficulty in diagonalizing a matrix with N dimensions grows approximately as N^2 in computer memory and N^3 in time. This means that calculations with high J require increasingly sophisticated diagonalization procedures and computational power. The computational demands of the final line list are further described in Section 5.2.

When the SYT line list began to be generated, supervision was required to ensure that crashed jobs were correctly re-calculated, while loosing as little information as possible. In the case of the energy eigenvalues, if the calculations crashed before completing they would have to be re-initiated completely, as the diagonalizing algorithms do not allow for a feasible way of resuming the calculations. In the case of the transition intensity calculations, resuming after crashing was possible, but not without multiple small overlaps of transition calculations, and consequently duplicates in the final line list. These were eventually all found and carefully removed, but it was such a painstaking process that steps were taken to automate it for the production of the SAlTY line list.

Chapter 4

Thermodynamic Data on Phosphine and Ammonia

The methods used to calculate partition functions by previous sources become increasingly inaccurate at high temperatures, where it is necessary to consider the contribution from very many energy levels. The creation of the SYT line list described in section 3 required the production of 5.6 million theoretical energy levels for phosphine derived from variational calculations. Given that this provided access to an unprecedented number of energy levels, it was considered of value to perform a thermodynamic data analysis on phosphine using an explicit summation of these energies. This procedure has been used successfully before to provide reliable high temperature partition functions and related thermodynamic properties [115, 64, 172, 12, 69]. Some of these partition sums have differed by orders of magnitude from prior estimates, (see, for example, Ref. [115]). Additionally, previous work on ammonia [185, 183], NH_3, had provided 7.5 million energy levels and, given that the two molecules share many similarities, it was opportune to study the thermodynamic properties of both simultaneously.

Nicholas Hesketh, a summer student from the Nuffield Foundation, began investigating the partition function and the specific heat capacity of ammonia as part of a summer research project. This project had limited scope but Mr Hesketh successfully used a selection of the energy levels to find a reasonable estimate for both properties and provided an excellent platform from where to start the work described in this chapter.

Ammonia and phosphine are both symmetric top molecules, but differ in the inversion motion which has, so far, only been observed for ammonia. This tunneling effect is further discussed in section 6.3. The two molecules are of great interest for a variety of terrestrial and extra-terrestrial studies and have already been found in many astrophysical objects and interstellar media.

Both molecules have been for a long time known to be abundant in Jupiter and Saturn [51, 135, 47, 30, 89, 159], and are expected to be present in extrasolar gas giants [15]. Ammonia spectra are thought to be the key for identifying the coldest class of brown dwarf stars, so-called Y-dwarfs [97, 146]. Detecting the presence of either molecule in environments with high temperatures requires not just line lists that are complete to those temperatures, but also a partition function that is valid at those temperatures.

Accurate thermodynamic data on these molecules is therefore of great importance. In particular, the partition function is necessary to establish the correct temperature dependence of spectral lines and their intensity, as given by 2.7, which has the following proportionality:

$$I(T_1) \propto I(T_2)\frac{e^{-E_i/T_1}}{e^{-E_i/T_2}}\frac{Q(T_1)}{Q(T_2)}, \tag{4.1}$$

where the I(T_1) is the intensity of the transition at the extrapolated temperature T=T_1, I(T_2) is the measured/calculated intensity of the transition at T=T_2, E_i is the lower state energy for the transition, and $Q(T_i)$ is the temperature dependent partition function.

As is further elaborated in Section 1, this temperature extrapolation is not straightforward to evaluate. This is partially due to the difficulty in assigning transitions to their lower energy, E_i, but also because the partition function is not always accurately provided for the required range of temperatures.

4.0.1 Previous Thermodynamic Data

Previous work provides partition functions for both phosphine and ammonia but with limitations. In 1988, Irwin [74] published a least squares polynomial log fit to the partition functions for molecules of interest for stellar atmospheric equations of state, including NH_3 and PH_3; his results are presented as valid from 1000 K to 6000 K, but are increasingly unreliable at the higher end of this temperature range. Gamache *et al.* [55] fit total internal partition sums (TIPS) to a polynomial expression (third-order in temperature) for most of the infrared absorbing molecules in the terrestrial atmosphere, including both NH_3 and PH_3,

in three temperature ranges reaching up to 3005 K. Fischer *et al.*[50] used an interpolation scheme to calculate TIPS in the temperature range 70-3000 K for all molecular species in the HITRAN [139, 140] database. The Cologne Database for Molecular Spectroscopy (CDMS) [112] provides partition function values for temperatures up to 300 K, but these are too low as it would appear that only the rotational contribution to the partition function is represented.

In 1968 Haar [63] presented thermodynamic tables for ammonia for temperatures between 50 K - 5000 K. The 1985 JANAF tables [32] present thermodynamic properties for both molecules for temperatures up to 6000 K but, again, these are increasingly unreliable at high temperatures and have been found to be partially incorrect [173]. In 1999, Lodders [92] updated the JANAF thermodynamic tables for some phosphorus molecules, including some corrections to the phosphine data, but did not significantly alter JANAF's results. Finally, Cheric [33] is an online resource which provides calculations for the heat capacity of ammonia and phosphine, but no temperature limits are provided for either molecule and all values for phosphine appear to be unphysical.

The results from the studies cited above are discussed and compared to those presented here in Section 4.2.

4.1 Method

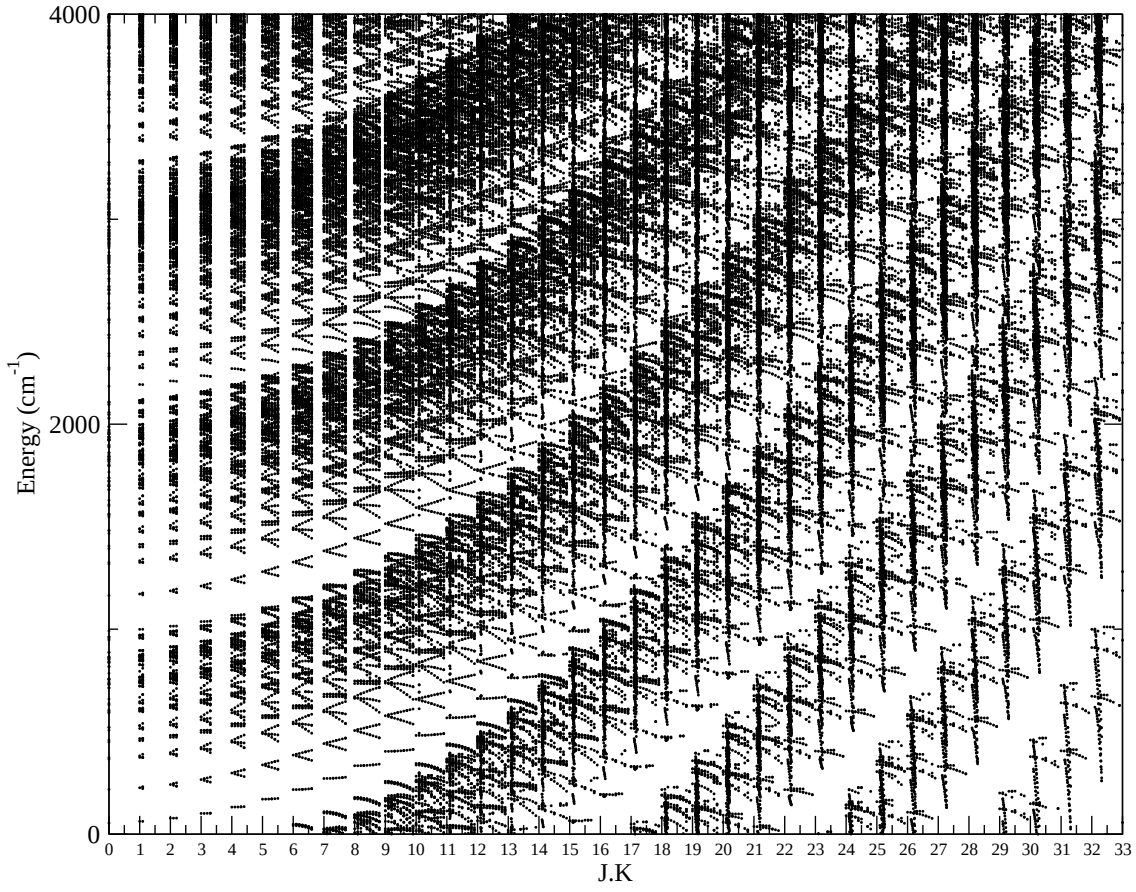
4.1.1 Energy level calculations

The process used here for calculating the total internal partition sums (TIPS) and related thermodynamic properties follows closely that used by Vidler et al [172] for water, which builds on methods described by Harris et al [64] and Martin et al [107]. It uses the explicit summation of theoretical rotation-vibration energy levels as a reliable method for producing highly accurate internal partition functions, even for high temperatures. Extension to the total partition functions is done using the ideal gas approximation.

The energy levels for both molecules were calculated using the program TROVE [193]. Both room temperature [185] and elevated temperature [183] line lists for ammonia have already been produced and are freely available online. The ammonia energy levels used here are from the high temperature line list, known as BYTe, as it is both more accurate and complete than its predecessor. The energy states from the phosphine room temperature line list, SYT, described in section 3, were used here.

Approximately 7.5 and 5.6 million energy levels were computed for ammonia and phosphine, respectively, part of which were used in the creation of the molecular line lists. An additional 145 million energy levels were computed for phosphine specifically for the thermodynamic calculations presented here, since the original 5.6 million did not allow for a satisfactory level of convergence. These were created with an accurate vibrational component but with the rotational contribution estimated by a rigid rotor approximation, and consequently with a much decreased level of accuracy. The rotational states are anchored to the vibrational states, so an accurate vibrational description means that the energy level clusters remain valid [194], even if the degradation of accuracy from the rotational component makes the energies within the cluster only approximate. Figure 4.1 illustrates this clustering.

Figure 4.1: Clustering of energy states with quantum numbers J and K. In the x-axis, the J,K values are represented as J.K (e.g J=33,K=29 is 33.29).



The collective contribution of the states to the partition function and related properties remains valid. Some of the calculated energy levels corresponded to quasi-bound states, but the calculations in this chapter were made excluding any energy levels over the dissociation

threshold of the molecules. The precise role of quasi-bound states in the partition function and related properties is still not fully understood but our results suggest that they are unlikely to make a significant contribution at temperatures of interest to ammonia and phosphine.

Nuclear Statistical Weights

The internal rotation-vibration functions of ammonia, with its small barrier to the planarity and distinct inversion tunnelling through the 'umbrella' mode, are best described by the $\mathcal{D}_{3h}$ molecular symmetry group. It spans six irreducible representations A_1', A_2', E', A_1'', A_2'', and E'' [25]. The molecular symmetry group of the rigid molecule PH_3, $\mathcal{C}_{3v}$(M), with its high barrier to the inversion, is isomorphic to the point group, C_{3v}, which spans three irreducible representations A_1, A_2, E. Combination of the internal ro-vibrational function with the nuclear spin functions is subject to Fermi-Dirac statistics: the hydrogen nuclei are fermions with spins of $1/2$ and thus the total nuclear-rotation-vibration function must be asymmetric with respect to the interchange of two protons. As a result the nuclear statistical weights, g_{ns}, are introduced in order to account for this property. For ammonia $g_{ns} = 0$, 12, 6, 0, 12 and 6 for the A_1', A_2', E', A_1'', A_2'', and E'' internal ro-vibrational eigenstates, respectively. The A_1' and A_1'' states have been given zero weights as these levels do not exist in nature. In the case of phosphine $g_{ns} = 8$ for all three symmetries A_1, A_2 and E. It is important to note that in this and other work within the ExoMol project [165] the 'physics' convention for the definition of statistical weights is followed. This convention includes all nuclear spin degeneracies in the total internal partition sum (TIPS), a convention also followed by HITRAN [55]. Conversely, the 'astrophysics' convention, followed for example by JANAF [32] and Irwin [74], normalize the TIPS to unity for each atom. This results in TIPS which differ by integer multiples associated with the degeneracy of the atomic nuclear spins; the statistical weight factors are $2 \times 2 \times 2 \times 3 = 24$ for ammonia and $2 \times 2 \times 2 \times 2 = 16$ for phosphine. Many results, such as line intensities, rely on ratios which are independent of this choice of convention as long as it is applied consistently. In other cases however, such as with entropy, the results differ by a constant shift. These issues are further discussed in chapter 4.2 and by Goldman et al [58].

4.1.2 Energy Thresholds

When calculating the partition function through the explicit summation of energy levels, the molecular symmetry group plays an important role, by helping to reduce the size of the problem and reach a higher temperature coverage for the data produced. In the explicit summation approach, the main bottleneck in calculating the partition function applicable for high temperatures is the high computational cost associated with obtaining energies with the high-lying rotation-vibration excitations required. For example, in the case of the BYTe line list for NH_3 [183], the energies from which are used here, the highest J was 41 and some severe energy thresholds were used ($E_{\max} = 18\ 000$ cm^{-1}). With these thresholds the applicable temperature range for BYTe was estimated to be around 1500 K. Increasing this threshold was prohibited by the size of the Hamiltonian matrices of the E-type symmetry to be diagonalized, which were larger than 100 000$\times$100 000 [183]. Similar thresholds were employed for the phophine SYT line list.

To overcome these threshold limitations, an approach for computing the partition functions based on the exclusive use of A-symmetry energies for the highest levels is considered. Variations of this approach have been tried in the past, in particular by Quack [136]. This technique requires significantly less computational resources because it does not require all of the E symmetry calculations, which are the most costly. It is described at length in the following section, 4.1.3.

Theoretical energy levels are still subject to systematic and approximation errors. However, variational nuclear motion calculations can yield many more of these energy levels than can be readily observed, guaranteeing the temperature-dependent completeness of the summations used to compute thermodynamical quantities such as the partition function. This completeness is particularly important for obtaining convergence of the sums at higher temperatures. Additionally, as it was discussed in section 4.1.1, when calculating partition function it is more important to obtain an accurate description of the clustering of the energy levels than the exact calculated value of each individual energy.

4.1.3 Calculation of the thermodynamic data

The internal partition function of a molecule in the ground singlet electronic state, $Q_{\rm int}$ (throughout this work also referred simply as Q), is given by:

$$Q = \sum_{J=0}^{J_{\max}} \Delta Q_J = \sum_i g_i(2J_i+1)\exp\left(-\frac{c_2\tilde{E}_i}{T}\right), \tag{4.2}$$

where $c_2 = hc/k = 1.438777$ cm·K is the second radiation constant, J_i is the rotational quantum number of the $i^{\rm th}$ state which has term value $\tilde{E}_i$ (in cm^{-1}) relative to the $J = 0$ zero-point-energy, and g_i is the nuclear spin degeneracy factor. The summation in Eq. (4.2) runs over all the bound ro-vibration energy levels of the system but, as the energy of the states increases, the exponential factor decreases towards zero and the summation of energy levels may be truncated without loss of accuracy [55] as the partition sum will be converged. For higher temperatures, more energies significantly contribute to the sum and the convergence is harder to achieve. The main issue here is convergence with rotational excitation, given by J; Eq. (4.2) has been given in a form which explicitly recognizes the contribution of each J up to some maximum value, $J_{\max}$.

From equation 4.2, further internal thermodynamic functions can be calculated. The specific heat capacity, C_p, the Helmholtz function, hcf, the Gibbs enthalpy function, gef, and the entropy, S:

$$C_p(T) = R\left[\frac{Q''}{Q} - \left(\frac{Q'}{Q}\right)^2\right], \tag{4.3}$$

$$hcf(T) = H(T) - H_0 = RT\frac{Q'}{Q} - H_0, \tag{4.4}$$

$$gef(T) = -\frac{[G(T) - H_0]}{T} = R\ln Q + \frac{H_0}{T}, \tag{4.5}$$

$$S(T) = R\frac{Q'}{Q} + R\ln Q, \tag{4.6}$$

where H_0 is the enthalpy at the reference temperature of 298.15 K and Q' and Q'' are, respectively, the first and second moments of the internal partition function defined as

$$Q' = T\frac{dQ}{dT} = \sum_i g_i(2J_i+1)\left(\frac{c_2\tilde{E}_i}{T}\right)\exp\left(-\frac{c_2\tilde{E}_i}{T}\right), \tag{4.7}$$

$$Q'' = T^2\frac{d^2Q}{dT^2} + 2\frac{dQ}{dT} = \sum_i g_i(2J_i+1)\left(\frac{c_2\tilde{E}_i}{T}\right)^2\exp\left(-\frac{c_2\tilde{E}_i}{T}\right), \tag{4.8}$$

which are also calculated by explicit sums running over the energy levels [107]. These are further explained elsewhere [64, 172, 107].

Partition Functions by Symmetry

Calculations of the energy levels, partition sums and other thermodynamic properties were performed for the major isotopologues of phosphine and ammonia, $^{31}PH_3$ and $^{14}NH_3$. The procedures and potential energy surfaces used to create the energy levels are described in chapter 3 for PH_3 and the line list article for NH_3 [183]. These line lists contain ro-vibrational term values covering states up to $J_{\max} = 33$ and $\tilde{E}_{\max} = 12\ 000$ cm^{-1} for phosphine and $J_{\max} = 41$ and $\tilde{E}_{\max} = 18\ 000$ cm^{-1} for ammonia. These thresholds impose important limitations on the temperature ranges these line lists can be applicable for as well as on their corresponding partition functions, which were $T = 300$ K and 1500 K, for PH_3 and NH_3, respectively. For the present work we have extended these sets of energy term values to achieve higher temperature coverage of the thermodynamic properties studied. The highest energy levels computed for the work presented here were 74 488.02 cm^{-1} and 74 761.4 cm^{-1} with a maximum J value of 100 and 45, for phosphine and ammonia, respectively. The thermodynamic calculations were then made using only energy levels up to dissociation (28 839.7 cm^{-1} and 41 051 cm^{-1} for phosphine and ammonia [106], respectively), but the consequences of including states above this threshold were investigated and are further discussed below.

The main bottleneck and the source of limitations in computing energy levels for both ammonia and phosphine is the size of the Hamiltonian matrices to be diagonalized while solving the corresponding ro-vibrational Schrödinger equations or, more specifically, the sizes of the E (or E' and E'' in case of NH_3) symmetry matrices. For a given value of J, the dimensions of the A-symmetry matrices are approximately half the dimensions of the E-symmetry matrices. In order to cope with the matrix size problem the line list calculations employed an iterative eigensolver PARPACK [86] for higher J values. The drawback of this procedure is that it is necessary to restrict the number of eigen-roots, which resulted in the energy threshold of 18 000 cm^{-1} for ammonia and 12 000 cm^{-1} for phosphine. For accurate evaluation of the partition function using Eq. (4.2) at high T more eigenvalues are needed, which can be achieved by the use of direct eigensolvers (e.g. DSYEV from the LAPACK library [9]). However, the eigenvectors are not required, which usually reduces the computational costs. More importantly, the final eigenvalues are not necessarily required as they are needed only as input for the statistical sum. Yurchenko *et al.* [182] explored the property of the matrix trace and performed the temperature averaging of different ammonia properties employing the expansion of the matrix exponent thus avoiding the eigen-problem.

In this work a similar approach was taken.

Consider the partial statistical sum

$$Q_J^\Gamma = \sum_i \exp\left(-\frac{c_2 \tilde{E}_i^{J,\Gamma}}{T}\right), \tag{4.9}$$

where J is the rotational angular momentum quantum number, Γ is one of the irreducible representations and i runs over all states with a given J and Γ. Then for high J the partial sums Q_J^Γ of all 1D irreducible representations (for example A_1 or A_2 in case of $\mathcal{C}_{3v}$(M)) are approximately equal and are half of that of the 2D representation (e.g. E, E' or E''). This property has previously been studied for H_3^+ [18], although it was not used in computing their high-T partition sum [115]. Note that Q_J^Γ does not include the total degeneracy factor $g_i = J_i(J_i+1)g_{\rm ns}$.

For $\mathcal{C}_{3v}$(M) (PH_3) at high temperatures we expect:

$$Q_J^E \approx 2Q_J^{A_1} \approx 2Q_J^{A_2} \approx Q_J^{A_1} + Q_J^{A_2} \approx Q_J^A \tag{4.10}$$

Similarly, for $\mathcal{D}_{3h}$ (NH_3):

$$Q_J^{E'} \approx Q_J^{E''} \approx 2Q_J^{A_1'} \approx 2Q_J^{A_2'} \approx 2Q_J^{A_1''} \approx 2Q_J^{A_2''} \approx Q_J^A \tag{4.11}$$

The conjecture from the equations above is based on the fact that for high J the density of the A_1, A_2 and E matrices is similar but the A_1 and A_2 symmetries have approximately half the number of energy states as the E symmetry. This was previously visible when calculating SYT, and Figure 3.4 in Chapter 3.4.1 shows that, after a J threshold, the number of E symmetry energy levels is approximately twice as high as that of either A symmetry.

Actual energy values for a given J are very similar for all symmetries. Then it follows that to evaluate the Q_J^Γ-contributions to the partition function at high J only A-symmetries are required. This is important because in the variational computations, calculations of the E-symmetry energies associate with high computational costs. Using equations 4.10 and 4.11 the E-symmetry energies can be dismissed without meaningful loss of accuracy, at least above some temperature-dependent J threshold, J'. The total internal partition function can be closely approximated using

$$Q_{\rm int} = \sum_{J=0}^{J'}(2J+1)(g_A Q_J^A + g_E Q_J^E) + \sum_{J=J'+1}^{J_{\max}}(g_A+g_E)(2J+1)Q_J^A, \tag{4.12}$$

where $Q_J^A = Q_J^{A_1} + Q_J^{A_2}$ and Q_J^E are the total reduced internal statistical sums for all A-type and E-type symmetries, respectively, and $g_A = g_{A_1} + g_{A_2}$ and g_E are the corresponding nuclear statistical weights. Here all energies of PH_3 and NH_3 are computed by direct diagonalization of the corresponding Hamiltonian matrices.

Approximation Limitations

Equation (4.12) was found to be invalid at low temperatures ($T \leq 86$ K for ammonia and $T \leq 54$ K for phosphine), for all available values of J'. Under this temperature, the partition function can be calculated with the standard partition function sums using both symmetries, with $g_A Q_A^J + g_E Q_E^J$, as both line lists are complete up to these temperatures. Above these threshold temperatures, $(g_A + g_E)Q_A^J - (g_A Q_A^J + g_E Q_E^J)$ rapidly goes to zero and eq. (4.12) can be used without significant loss of accuracy. This approximation is valid with better than 0.05% accuracy for $J' = 18$ for both molecules but, to ensure the best possible results, this approach should only be used in the absence of *E*-symmetry energy levels, at which point an explicit summation over the energies of all symmetries is impossible.

To ensure that equation 4.12 was used only where valid and where the relevant E-symmetry energies were absent, the TIPS were calculated using a hybrid approach. Since complete data for all symmetries is available for $J \leq 20$ for phosphine and $J \leq 23$ for ammonia, the standard partition function sums were performed up to these values of J. For $24 \leq J \leq 40$ (for ammonia) and $21 \leq J \leq 33$ (for phosphine) the *E*-symmetry energy levels were still considered but, due to their energy truncation at 18 000 cm^{-1} (ammonia) and 12 000 cm^{-1} (phosphine), the standard sum was used under these energy thresholds, and exclusively the *A*-symmetries over them, with $(g_A + g_E)Q_A^J$. For $J \geq 40$ (ammonia) and $J \geq 33$ (phosphine), due to the absence of any E symmetry data, only the A symmetry energy levels were considered. In the case of the phosphine molecule, given the poor convergence achieved by the $J \leq 33$ data alone, the additional approximate data discussed in chapter 4.1.1 was incorporated in the calculations. All states with $J \geq 34$ from this data were used as well as those with $J \leq 33$ and energy values over the threshold of the original data.

Figure 4.2 shows the difference in the values of the calculated TIPS when using the standard partition function sums using both symmetries, $g_A Q_A^J + g_E Q_E^J$, the approximation using A symmetries only, $(g_A + g_E)Q_A^J$, and the hybrid approach described above. The accuracy of Eq. (4.12) means that the hybrid approach makes only a negligible difference in the val-

ues and convergence of the partition function when compared to using only A symmetries. However, this small improvement in accuracy is amplified when calculating the other thermodynamic properties, in particular the heat capacity, as it depends on both the first and the second moments of the partition function.

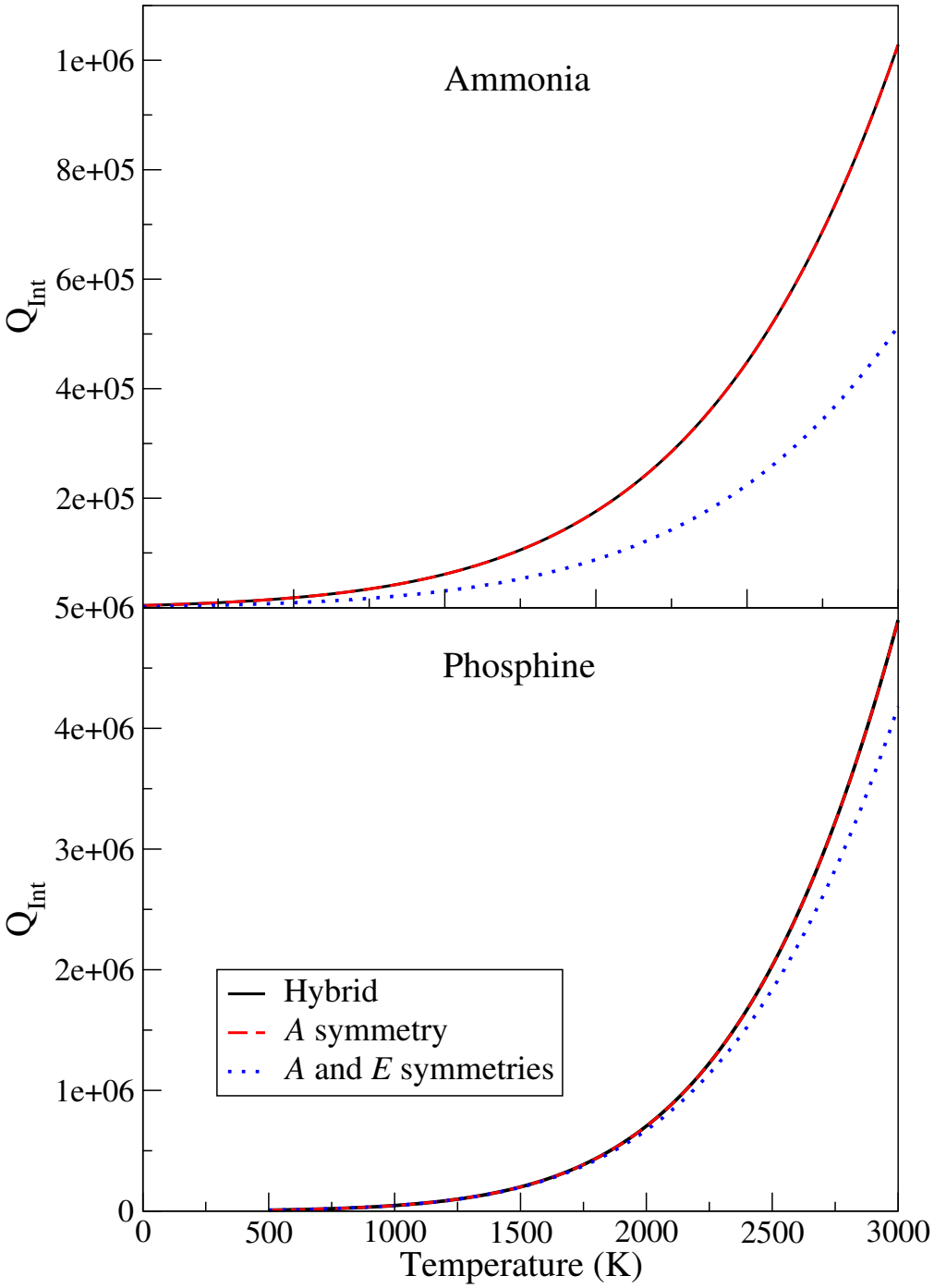


Figure 4.2: Comparison of the partition function values calculated with three different approaches. The A symmetry only, and hybrid approaches produce very similar results and at this scale are indistinguishable.

The partition functions calculated here are the total internal partition functions. In order to extend this to the total partition function the ideal gas approximation was used. In this approximation, the internal (Eqs. 4.3 to 4.6), translational and total thermodynamic properties relate to each other as follows:

$$C_p^{\mathrm{Total}}(T) = C_p^{\mathrm{Int}}(T) + \frac{5}{2}, \tag{4.13}$$

$$hcf^{\mathrm{Total}}(T) = hcf^{\mathrm{Int}}(T) + \frac{5RT}{2}, \tag{4.14}$$

$$gef^{\mathrm{Total}}(T) = gef^{\mathrm{Int}}(T) + R\ln\left(\frac{kT}{p_0\Lambda^3}\right), \tag{4.15}$$

$$S^{\mathrm{Total}}(T) = S^{\mathrm{Int}}(T) + \frac{5R}{2} + R\ln\left(\frac{kT}{p_0\Lambda^3}\right), \tag{4.16}$$

where R is the gas constant, k is the Boltzmann constant, p_0 is the standard state pressure, 0.1 MPa, and Λ is the thermal de Broglie wavelength, $h/\sqrt{2\pi k M_r u T}$, with $M_r u$ being the molecular mass in kg, which equals the molecular weight, M_r, multiplied by $10^3 M_u N_A^{-1}$ (molar mass constant and Avogadro number). The de Broglie wavelength, Λ is directly related to the translational component of the partition function, with $\mathrm{Q_{trans}} = \Lambda^{-3}$, or $(2\pi k M_r u T/h^2)^{\frac{3}{2}}$ in the ideal gas approximation.

There were some difficulties in obtaining the correct form of the translational component of the thermodynamic properties, as the literature on the matter was frequently inconsistent. Equations 4.13 to 4.16 are a rearrangement and correction to those published alongside the JANAF thermochemical tables [32]. Corrections are made to the ambiguous treatment of the molecular mass and its units, as well as a typo in the original where N should read N_A.

4.2 Results

The calculated values for the total internal partition function, specific heat capacity, Gibbs enthalpy function, Helmholtz function and entropy can be seen, respectively, in figures 4.3, 4.4, 4.5, 4.6, 4.7, for both ammonia and phosphine. These results are compared to previously determined values, and discussed individually in this section. Although the maximum temperature considered is 6000 K, there is no expectation that either molecule can be found in any significant abundance at temperatures above 4000 K.

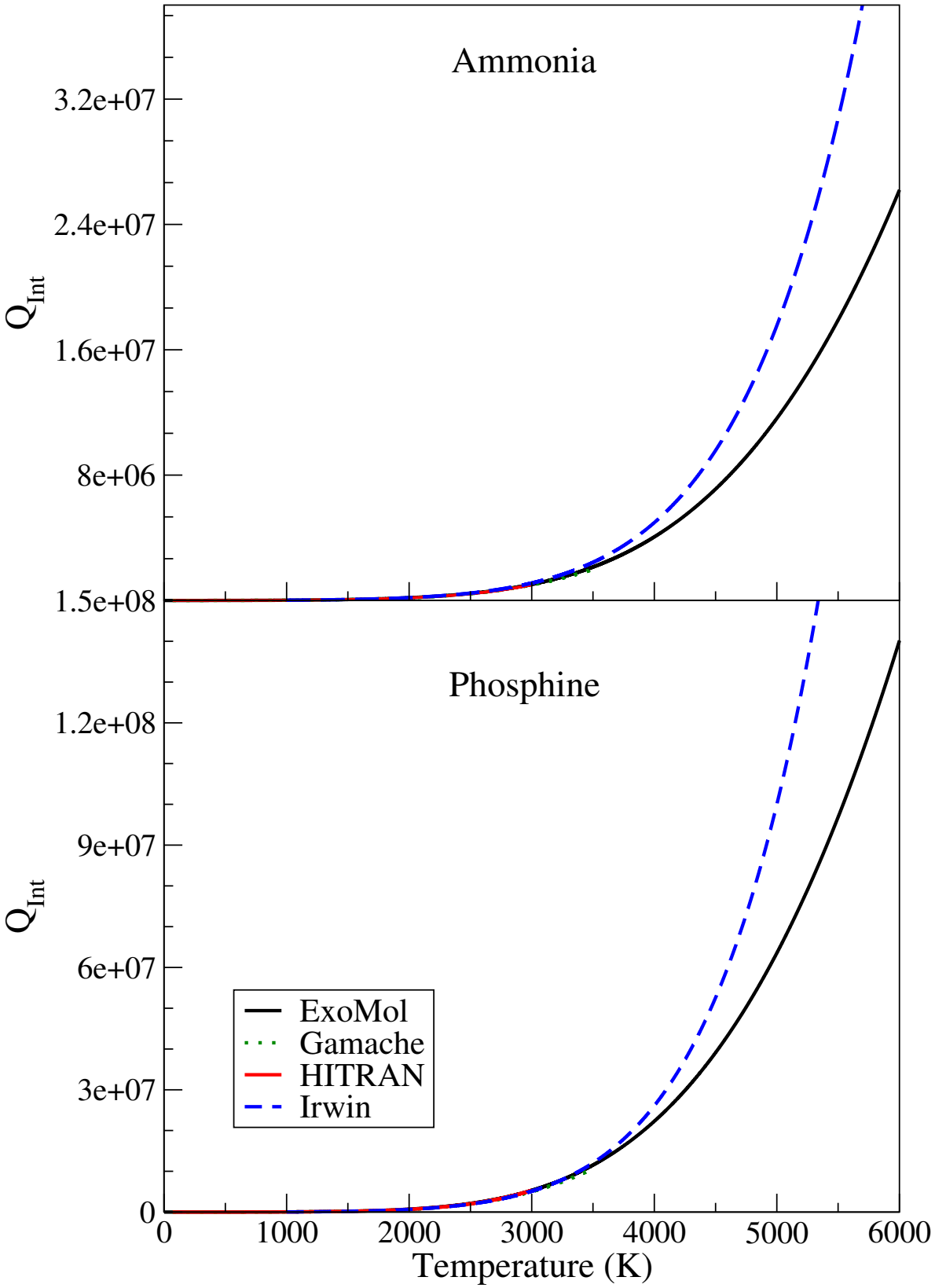


Figure 4.3: Comparison of the partition function values calculated here with those from Irwin [74], Gamache [55] and HITRAN [140, 50], for ammonia and phosphine. The HITRAN curve ends at 3000 K since that is the limit of the data provided.

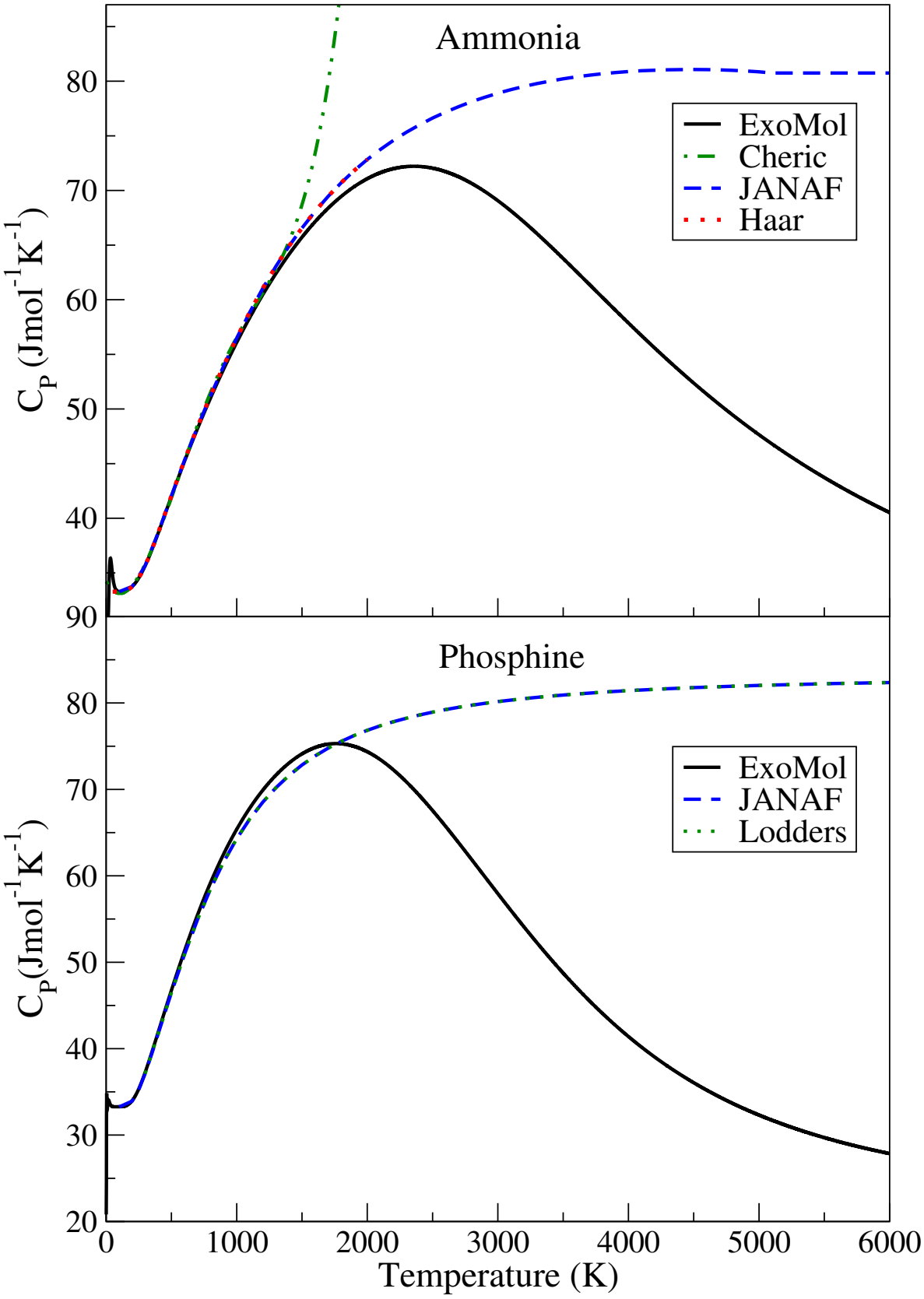


Figure 4.4: Comparison of the heat capacity function values calculated here with those from JANAF [32], Lodders [92], Haar [63] and Cheric [33], for ammonia and phosphine. Lodders' data is a revised version of that of the JANAF thermochemical tables, so the disagreement between them is minimal and the Lodders data remains mostly invisible at the scale shown here.

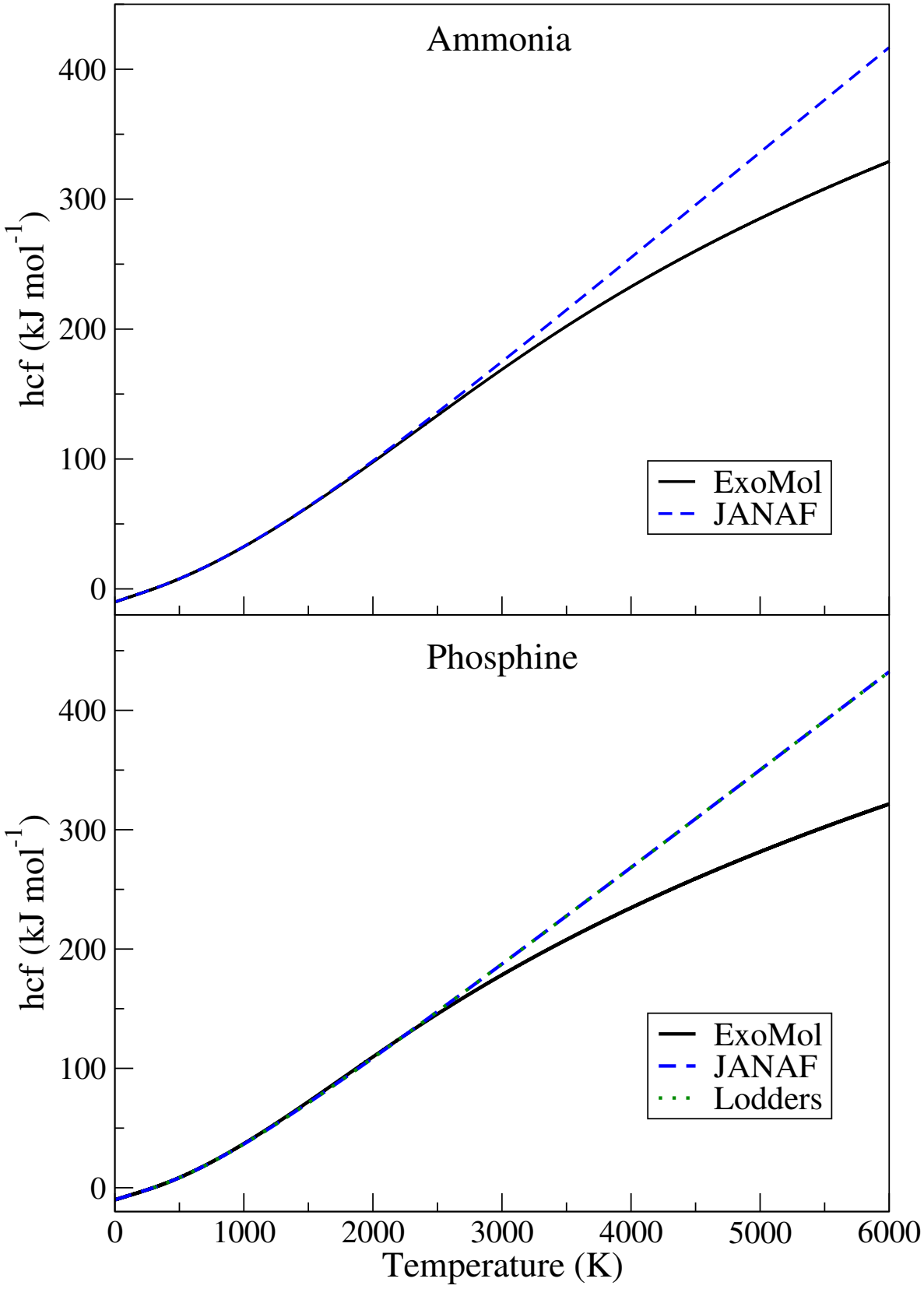


Figure 4.5: Comparison of the Helmholtz function values calculated here with those from JANAF [32], Lodders [92], for ammonia and phosphine.

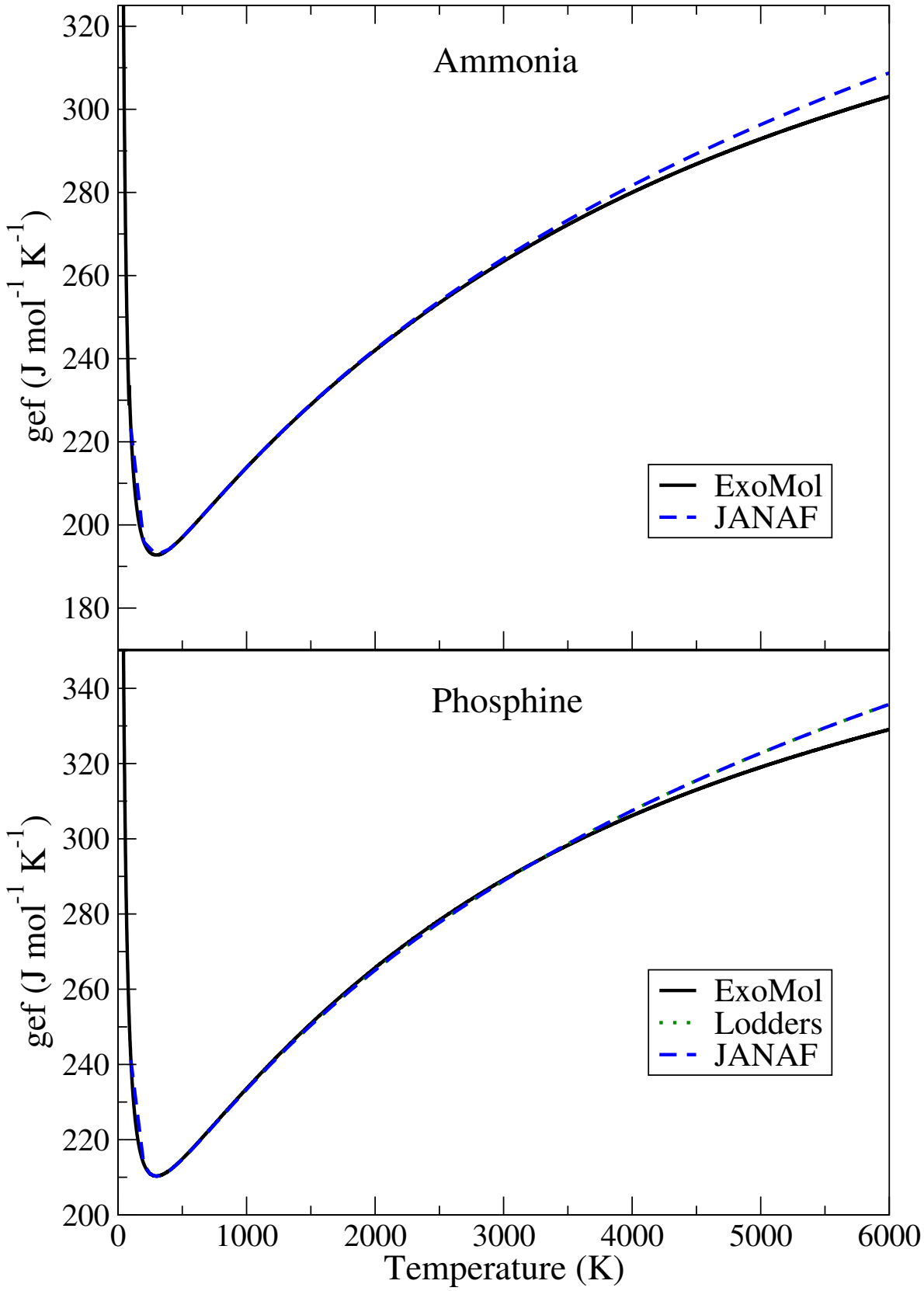


Figure 4.6: Comparison of the Gibbs free energy values calculated here with those from JANAF [32], Lodders [92], Haar [63] and Cheric [33], for ammonia and phosphine.

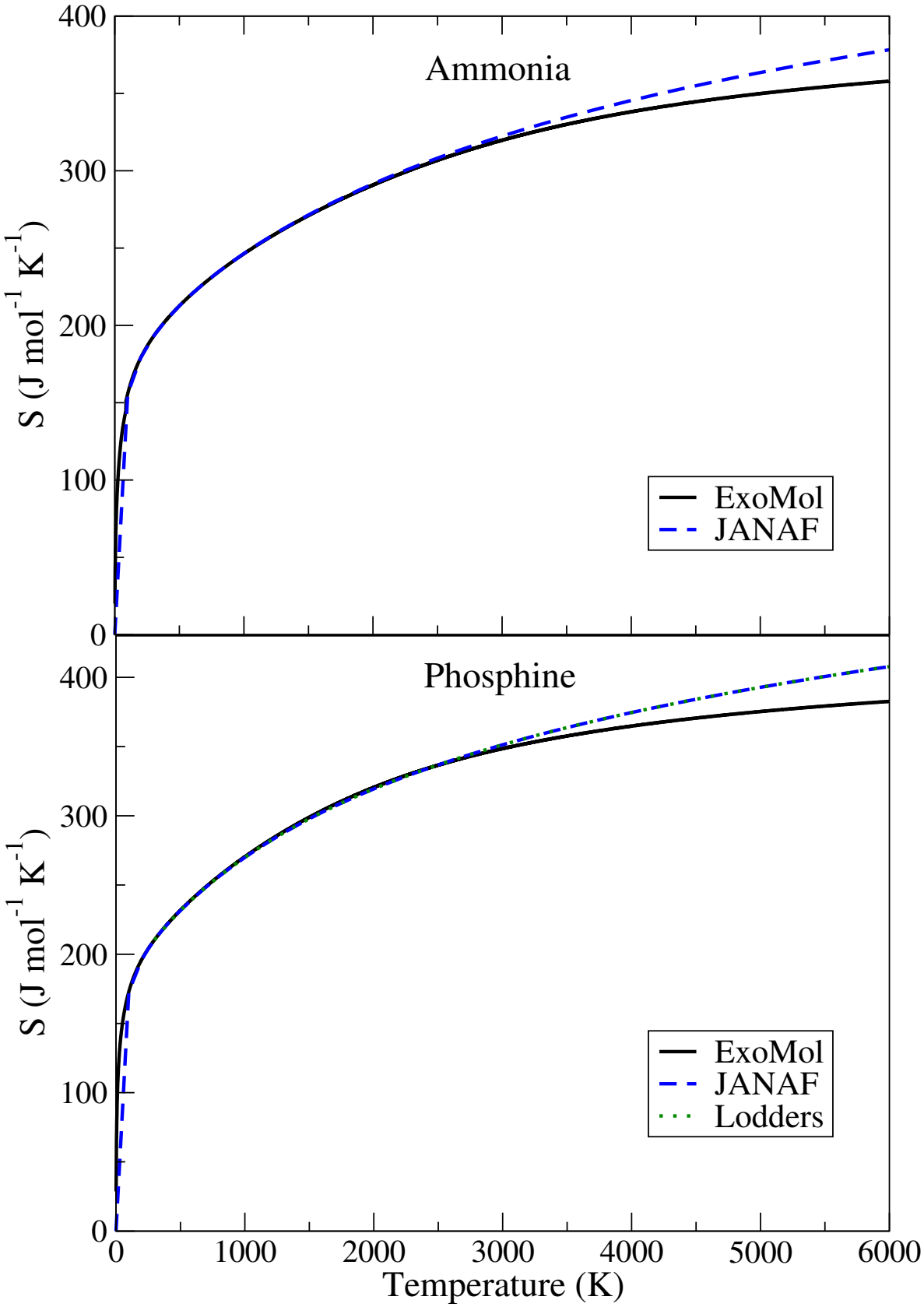


Figure 4.7: Comparison of the entropy values calculated here with that of those from JANAF [32], Lodders [92], for ammonia and phosphine.

The higher the temperature of the system being studied, the more energy levels will be expected to be populated, with higher values of J. This translates to an increase in difficulty when trying to calculate the partition function by explicit summation as the temperature rises. The rotational contribution to the partition function, ΔQ_J, should peak at a particular value of J and then decrease as the value of Q_{int} converges. Figure 4.8 shows the convergence of the value of the calculated partition function for a variety of temperatures for phosphine and ammonia. This was calculated by measuring the difference in contribution to the partition function made by consecutive J values, as a fraction of the total partition function, Q_{Int}. The partition functions calculated here are converged ($\leq 0.1\%$) for temperatures below 3220 K for ammonia, and up to the maximum temperature considered, 5000 K, for phosphine. In the convergence plot for phosphine, a small discontinuity can be seen at $J = 21$ and another at $J = 41$. This arises from the change of model employed in the calculation of energy levels and in the application of the hybrid approach to calculating the partition functions, and has an almost negligible impact on the accuracy of the final values.

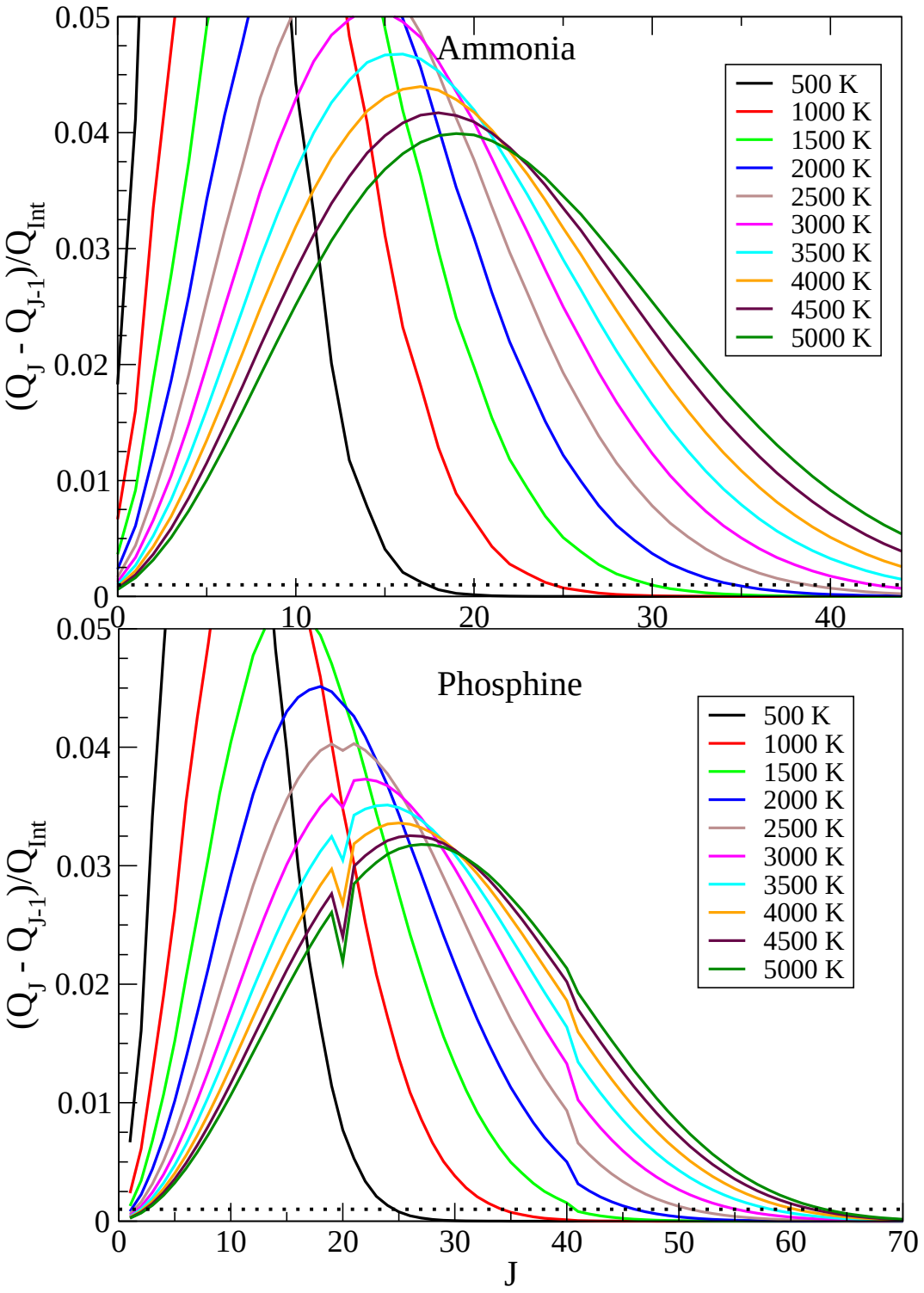


Figure 4.8: Convergence of the partition function, Q_{Int}, with respect to J, for phosphine and ammonia. Q_{Int} for all temperatures displayed is converged up to 1%. The dotted line represents the 0.1% convergence threshold. As described elsewhere, the discontinuities in the plot for phosphine arise from the change of model used for the calculation of energy levels with $J \geq 41$, and the hybrid approach used to calculate the partition functions.

4.2.1 Partition Functions

Fig. 4.3 compares the partition function calculated here with previous results from the literature [50, 140, 152]. All values compare well at low temperatures but then diverge when as T increases. Fischer *et al.* [50] considered the temperature range 70 - 3000 K and calculated TIPS by Lagrange 4-point interpolation. No polynomial fits are provided but data tables are available from the HITRAN database [140]. These data were retrieved from the VAMDC portal HITRAN-online [70]. The most recent HITRAN values of the partition function for 296 K are 1725.2 for ammonia and 3248.6 for phosphine [152]. These compare well with the present partition function values of 1725.24 and 3249.46 for ammonia and phosphine, respectively.

Gamache *et al.* [55] employed a polynomial equation of the format

$$Q(T) = a + bT + cT^2 + dT^3 \tag{4.17}$$

for the three separate temperature ranges, 70 - 500, 500 - 1500 and 1500 - 3005 K. These fits are made to the 2000 HITRAN database partition sums using the corrections mentioned by Goldman *et al.*[58] and match their calculated partition sums for both NH_3 and PH_3 within 1%.

Irwin [74] fits polynomial equations on a 100 K grid of temperatures ranging from 1000 to 6000 K using the fitting form

$$\log Q(T) = \sum_i a_i \left[\log\left(\frac{5040}{T}\right)\right]^i . \tag{4.18}$$

The maximum relative errors in Q over the fitting range are stated as between 10^{-2} – 10^{-1} for ammonia and 10^{-1} – 0.5 for phosphine. As previously discussed, Irwin's data differs from the work presented here by a factor of 24, for ammonia, and 16, for phosphine, due to the use of different statistical weights.

The partition functions calculated here can also be represented by a 8^{th} order polynomial

$$\log Q(T) = \sum_i a_i \left[\log(T)\right]^i \tag{4.19}$$

with coefficients shown on table 4.1. These fits reproduce the values presented here with accuracy of better than 0.4% for temperatures under 5000 K, but become increasingly unreliable when applied to higher temperatures. The results for the partition function as well as the other thermodynamic functions, are all provided, in 1 K intervals, as supplementary material to the

article associated with this chapter. The 1 K intervals should be more than sufficient for an accurate interpolation to any intermediate temperatures. Additionally, the partition function files contain its cumulative value with growing J, and its first and second moment for each temperature.

Table 4.1: Coefficients of the polynomial fit to our partition functions, $Q(T)$, see Eq. (4.19).

	Ammonia	Phosphine
a_0	0.5992996	0.9140408
a_1	-1.472838	-0.8504775
a_2	11.74190	3.029527
a_3	-23.25519	-3.209923
a_4	21.53000	2.137178
a_5	-10.42394	-0.6765168
a_6	2.727102	0.03679531
a_7	-0.3635381	0.02526486
a_8	0.01926157	-0.00382411

4.2.2 Heat Capacity

The heat capacity, C_p, is harder to converge due to its dependence on both the first and the second moments of the partition function. As can be seen by the corresponding convergence plots in Figure 4.9, the heat capacity is only fully converged ($\leq 0.1\%$) for temperatures below 2250 K for ammonia but is fully converged up to the maximum temperature considered, 5000 K, in the case of phosphine.

Using Eq. 4.13, the total heat capacity functions were compared to that of JANAF [32] and Lodders [92] for phosphine, and JANAF, Haar [63] and Cheric (*www.cheric.org*) for ammonia. Cheric provides a polynomial fit for both molecules, but the coefficients for phosphine create an extremely unreliable curve and have not been included in this comparison. All other comparisons are shown in Figure 4.4.

As before with the partition function, there is a very good agreement between the data at low temperatures, but a divergence as the temperature increases. Our value of the heat capacity at the reference value of 298.15 K is 37.2264 $\mathrm{Jmol^{-1}K^{-1}}$ for phosphine, which compares well to that of JANAF (37.102) and Lodders (37.101). Ammonia compares even better, with 35.638 $\mathrm{Jmol^{-1}K^{-1}}$, compared to that of JANAF (35.652), Haar (35.654) and Cheric (35.782).

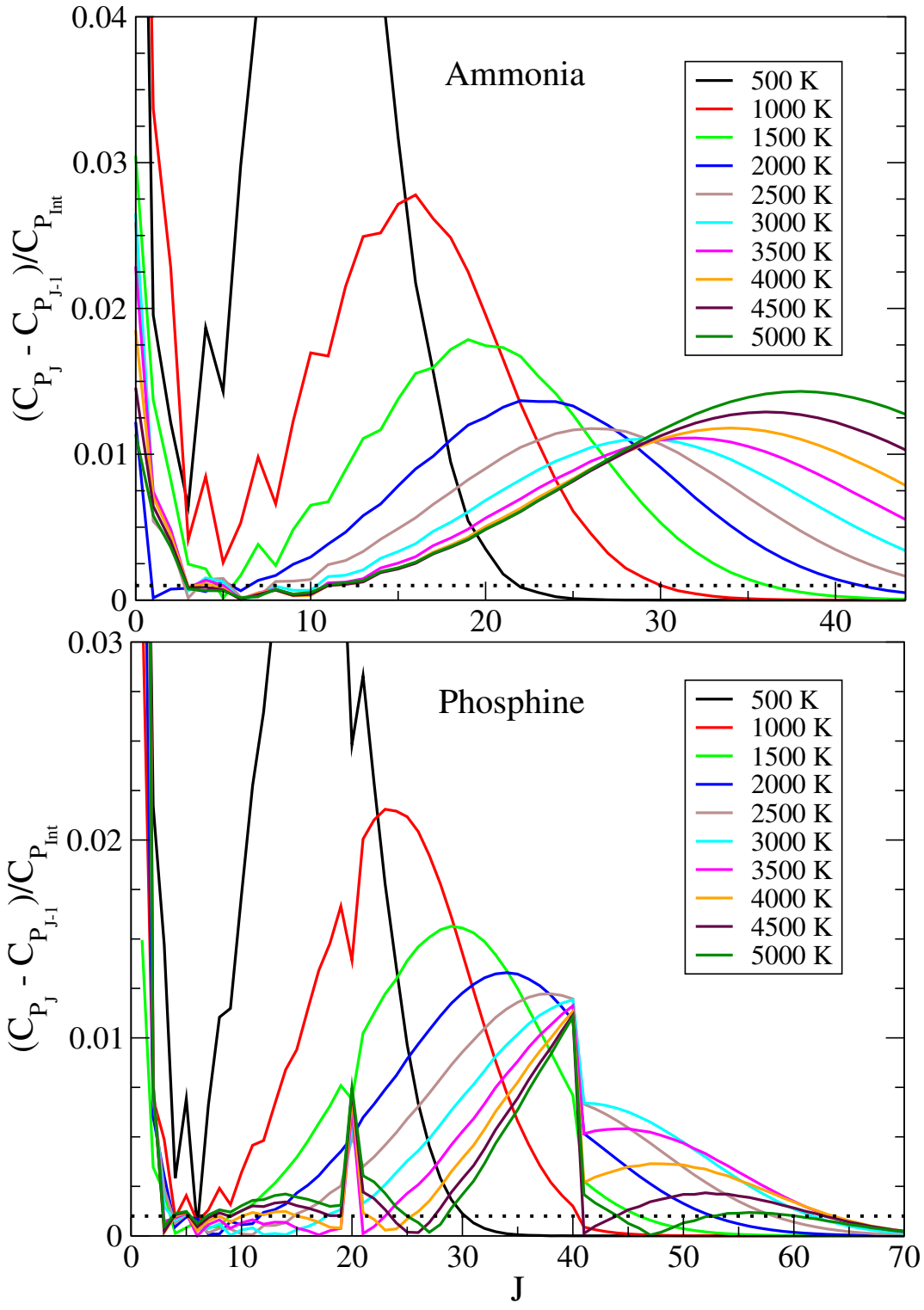


Figure 4.9: Convergence of the heat capacity function, C_{p}, with respect to J, for phosphine and ammonia. Q_{Int} for all displayed temperatures is converged up to 1%. The dotted line represents the 0.1% convergence threshold. As described above, the discontinuity in the plot for phosphine arises from the change of model used for the calculation of energy levels with $J \geq 41$.

Cheric's data is provided with no temperature limit, but it is probably only valid for low temperatures. Results from the other sources also become increasingly unreliable as temperatures rise. The primary source of thermodynamic species data are the JANAF tables. The JANAF data on ammonia and phosphine is presented as covering temperatures up to 6000 K and are based on the use of simplified semi-empirical models. One of the largest sources of error is that these models do not take into account the effect of the dissociation of a molecule, which has previously been shown to be important [172]. This is the cause for the large discrepancy between our predictions and the results found in the literature, for all presented thermodynamic properties but, as can be seen in figure 4.4, this is particularly clear in the heat capacity calculations.

To demonstrate this, a simplified model based on a rigid-rotor harmonic oscillator approximation was used to make a rough evaluation of the evolution of the partition and the heat capacity functions. It was found that the results follow the shape of the curve computed here when limited by the molecule's dissociation energy, and that predicted by JANAF when the energy levels are taken to infinity. This confirms that the JANAF data significantly overestimates the partition functions for these molecules at high temperature, which has a detrimental effect all thermodynamic properties presented for temperatures above 1500 K.

4.2.3 Helmholtz, Entropy and Gibbs Free Energy Functions

The total Helmholtz function, hcf, is calculated using Eqs. (4.5) and (4.15) and H_0, the enthalpy at the reference temperature of 298.15 K. H_0 is calculated to be 10.1381 kJ mol^{-1} for phosphine and 10.0440 kJ mol^{-1} for ammonia, which compares well to the JANAF values of 10.136 kJ mol^{-1} and 10.045 kJ mol^{-1}. A comparative overview is shown in figure 4.5.

Other than near room temperatures, the Lodders data is indistinguishable from that of JANAF. This is because only temperatures near 296 K were affected by the revision due to Lodders.

As before, a factor of 24 (ammonia) and 16 (phosphine) on the partition function was required for the entropy and Gibbs free energy to be comparable to the JANAF literature. This does not affect the Helmholtz function or the heat capacity function since these rely on the ratio of the partition function with its first moment, Q'. A comparative overview is shown in figures 4.6 and 4.7.

The JANAF data has only 5 data points for temperatures under 300 K, and none between 0 and 100 K, so it is difficult to compare results in this region using the shape of the curves. This is the cause for the large discrepancy at low temperatures between our values and those from JANAF in figures 4.6 and 4.7.

4.3 Discussion

High accuracy theoretical data has been used to calculate the partition function and related thermodynamical properties of ammonia and phosphine, as a function of temperature. For $J \geq 41$ (ammonia) and $J \geq 34$ (phosphine), only the A symmetry energy levels were considered, taking advantage of the fact that the E symmetry contribution to Q can be estimated using only the A-symmetry contribution, for particular thresholds of T and J'.

The partition function, specific heat capacity, the Gibbs enthalpy function, the Helmholtz function and the entropy are all represented as internal functions, without translational components. The total functions can be obtained using, for example, the methods described in Section 4.1.

Analysis of JANAF's thermochemical data suggests that the JANAF data significantly overestimates the partition functions for these molecules at high temperature, which has a detrimental effect on other thermodynamic properties for temperatures above 1500 K. This appears to be due to the an incorrect estimation or absence of consideration for the dissociation energy.

The values for the partition function agree well with JANAF and other sources for low temperatures and, although only fully converged ($\leq 0.1\%$) for temperatures below 3000 K for ammonia (phosphine calculations were fully converged for all temperatures considered), should be reasonably accurate for temperatures below 5000 K for both molecules. The results for the heat capacity function are only fully converged for temperatures below 2000 K for ammonia, but should still give reasonable estimations of the heat capacity function values for temperature below 3000 K. All other thermodynamic properties are expected to be reasonably accurate up to 3000 K.

Chapter 5

Hot Line List

The room temperature line list described in chapter 3, SYT, is a preliminary linelist. Although SYT contains many orders of magnitude more lines than any previous phosphine line list (137 million transitions between 5.6 million energy levels), it was only designed to be accurate for temperatures below 300 K, making it unsuitable for most astronomical studies. This chapter sets out to describe the creation of a comprehensive hot line list calculated for $^{31}PH_3$ in its ground electronic state, henceforth referred to as SAlTY (**S**ousa-Silva, **Al**-Refaie, **T**ennyson, **Y**urchenko). SAlTY is the first line list suitable for modelling phosphine spectra in environments up to 1500 K and contains almost 16.8 billion transitions between 7.5 million energy levels. The completeness and coverage of the SAlTY line list makes it particularly suitable for studies of non-LTE environments. In these environments even extremely weak lines can play an important role in the shape of a spectrum. This is the reason behind including transitions with extremely weak absorptions in SAlTY.

This hot phosphine line list required diagonalizatoin of matrix sizes and transition calculations that pushed the limits of existing computational power and memory, even with extensive parallelisation of the problem. The high performance computing centres used were subject to long queues and CPU quotas, and crashed jobs had to be studied and re-submitted hundreds of times. Consequently, the creation of the SAlTY line list took the best part of a year.

5.1 Overview of the SAlTY Line List

SAlTY is a catalogue of transitions, each characterised by a frequency, its lower and upper energy level, Einstein coefficient and quantum numbers. Together these fully describe the spectrum of the phosphine molecule within the frequency range 0 – 10 000 cm^{-1}, or wavelengths longer than 1 μm. It contains 16 803 703 395 transitions between 7 480 690 energy levels below 18 000 cm^{-1}, with rotational quantum number J values up to 46. The highest energy state considered is 18 000 cm^{-1} above the zero-point energy (5 213.9280 cm^{-1}) for phosphine, as the intensity of transitions to higher energy levels is too weak to be relevant to the PH_3 spectrum, even at temperatures of 1500 K. With a highest energy state of 18 000 cm^{-1}, the highest lower energy state considered is 8 000 cm^{-1}, to ensure that the line list is complete within the frequency range 0 – 10 000 cm^{-1}.

All phosphine transitions are within the ground electronic state, since the first excited electronic state is above the dissociation limit of the molecule.

The strongest SAlTY transition has an Einstein-A coefficient of $89.1\ s^{-1}$ while the weakest lines go down to $10^{-48}\ s^{-1}$. In non-LTE environments even extremely weak lines can play an important role in the shape of a spectrum. This is the reason behind including transitions with extremely weak absorptions in SAlTY. Consequences of dismissing even extremely weak lines are explored in chapters 2.2.1 and 6.1.2.

Similarly to the SYT format, described in chapter 3, The final line list is presented in the ExoMol format [164], with a transition file ordered in increasing transition frequency and an energy file. The former contains a description of each transition by its upper (final) and lower (initial) energy level reference numbers (f and i), as well as the electric dipole transition probability represented as a Einstein coefficient A_{if} in s^{-1}. Using this information and a temperature-dependent partition function, Q(T), the line intensity of each transition can be calculated for any given temperature. The latter connects each index with the description of the corresponding energy level. Each energy level is described by the quantum numbers associated with the molecular group symmetry for PH_3, C_{3v}(M)[25], and total angular momentum J. The quantum numbers for XY_3 molecules are quite complicated and have recently been discussed in detail for ammonia by [44]. Our PH_3 quantum numbers follow the same general principles but are somewhat simpler since we neglect the possibility of a tunnelling mode.

Each energy level is described by the same quantum numbers used for SYT :

$$n_1, n_2, n_3, n_4, L_3, L_4, L, \Gamma_{\rm vib}, J, K, \Gamma_{\rm rot}, \Gamma_{\rm tot}, \tag{5.1}$$

described in chapter 3.3.7. These twelve quantum numbers reduce ambiguity to the assignment of the energy levels. However, only J and $\Gamma_{\rm vib}$ are rigorous; at higher energies, energy states cannot necessarily be assigned unambiguous quantum labels. Apart from the quantum numbers above, the largest eigen-coefficient used to produce the theoretical assignment is also provided (see, for example, [190]).

Since PH_3 transitions obey the strict selection rules $A_1 \leftrightarrow A_2 \quad , E \leftrightarrow E$, and $\Delta J = 0, \pm 1$, no higher J values were considered because $J = 45$ is the highest value of J for which there are eigenvalues exist below 8 000 cm^{-1}, which is the highest lower energy threshold used in SAlTY.

Excerpts from the energy and transition files are given in tables 5.1 and 5.3, respectively. The columns for the energy file are described in table 5.2. As per the ExoMol convention, these are named SAlTY.transitions and SAlTY.states. The complete line list is freely available and can be downloaded from the Strasbourg data centre, CDS, via ftp://cdsarc.u-strasbg.fr/pub/cats/J/MNRAS/ or from the ExoMol website, www.exomol.com. The website also offers the opportunity to download PH_3 cross sections [71], for a variety of resolutions, temperatures and range.

Table 5.1: Extract from the SAlTY Energy file.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
N	$\tilde{E}$	$g_{\rm tot}$	J	$\Gamma_{\rm tot}$	K	$\Gamma_{\rm rot}$	L	n_1	n_2	n_3	n_4	L_3	L_4	$\Gamma_{\rm vib}$	$\lvert C_i\rvert^2$	s_1	s_2	s_3	b_4	b_5	b_6
4772	18117.126945	8	0	2	0	1	3	0	3	6	1	2	1	2	1.00	1	3	2	3	0	1
4773	18141.236226	8	0	2	0	1	0	7	1	1	1	1	1	2	1.00	0	0	8	0	0	2
4774	18144.653263	8	0	2	0	1	3	0	4	5	2	5	2	2	1.00	1	2	2	2	4	0
4775	18180.965326	8	0	2	0	1	3	1	5	4	1	2	1	2	1.00	0	3	2	0	3	3
4776	18218.349734	8	0	2	0	1	0	0	4	4	4	4	4	2	1.00	1	2	1	0	5	3
4777	18287.993223	8	0	2	0	1	3	7	0	1	2	1	2	2	1.00	0	0	8	0	1	1
4778	18329.569862	8	0	2	0	1	6	0	6	1	9	1	5	2	1.00	0	0	1	0	15	0
4779	18393.321746	8	0	2	0	1	3	0	5	1	10	1	4	2	1.00	0	0	1	0	15	0
4780	18453.246434	8	0	2	0	1	3	0	5	1	10	1	4	2	1.00	0	0	1	0	15	0
4781	18506.447815	8	0	2	0	1	6	0	4	1	11	1	7	2	1.00	0	0	1	15	0	0
4782	18516.395845	8	0	2	0	1	3	6	0	3	0	3	0	2	1.00	8	0	1	0	0	0
4783	18548.610530	8	0	2	0	1	3	0	3	1	12	1	2	2	1.00	0	0	1	15	0	0
4784	18649.304702	8	0	2	0	1	6	0	4	1	11	1	7	2	1.00	0	0	1	15	0	0
4785	18725.969827	8	0	2	0	1	0	8	0	1	1	1	1	2	1.00	9	0	0	0	0	1
4786	19353.229098	8	0	2	0	1	6	0	7	1	9	1	7	2	1.00	0	0	1	16	0	0
4787	19493.996332	8	0	2	0	1	9	0	6	1	10	1	8	2	1.00	0	0	1	0	0	16
4788	19745.979241	8	0	2	0	1	0	8	1	1	1	1	1	2	1.00	9	0	0	0	0	2
4789	19894.271775	8	0	2	0	1	3	8	0	1	2	1	2	2	1.00	9	0	0	0	1	1
4790	20126.886789	8	0	2	0	1	3	7	0	3	0	3	0	2	1.00	9	0	1	0	0	0
4791	1118.304691	8	0	3	0	1	1	0	0	0	1	0	1	3	1.00	0	0	0	0	0	1
4792	2108.150565	8	0	3	0	1	1	0	1	0	1	0	1	3	1.00	0	0	0	0	0	2
4793	2234.920254	8	0	3	0	1	2	0	0	0	2	0	2	3	1.00	0	0	0	0	1	1
4794	2326.870042	8	0	3	0	1	1	0	0	1	0	1	0	3	1.00	0	0	1	0	0	0
4795	3085.609104	8	0	3	0	1	1	0	2	0	1	0	1	3	1.00	0	0	0	0	2	1
4796	3222.494320	8	0	3	0	1	2	0	1	0	2	0	2	3	1.00	0	0	0	3	0	0
4797	3311.958593	8	0	3	0	1	1	0	1	1	0	1	0	3	1.00	0	0	1	0	0	1
4798	3333.494686	8	0	3	0	1	1	0	0	0	3	0	1	3	1.00	0	0	0	2	1	0
4799	3424.626917	8	0	3	0	1	1	1	0	0	1	0	1	3	1.00	1	0	0	0	0	1
4800	3435.624836	8	0	3	0	1	2	0	0	1	1	1	1	3	1.00	0	0	1	0	0	1
4801	4050.520058	8	0	3	0	1	1	0	3	0	1	0	1	3	1.00	0	0	0	3	0	1
4802	4196.913916	8	0	3	0	1	2	0	2	0	2	0	2	3	1.00	0	0	0	0	2	2
4803	4283.755426	8	0	3	0	1	1	0	2	1	0	1	0	3	1.00	0	0	1	0	1	1
4804	4319.935065	8	0	3	0	1	1	0	1	0	3	0	1	3	1.00	0	0	0	4	0	0
4805	4408.422285	8	0	3	0	1	1	1	1	0	1	0	1	3	1.00	1	0	0	0	0	2
4806	4418.431045	8	0	3	0	1	2	0	1	1	1	1	1	3	1.00	0	0	1	0	0	2
4807	4438.012026	8	0	3	0	1	2	0	0	0	4	0	2	3	1.00	0	0	0	0	2	2

Table 5.2: Legend for the energy file, SAlTY.states.

Column	Notation	
1	N	Energy level reference number (row)
2	$\tilde{E}$	Term value (in cm^{-1})
3	g_{tot}	Total degeneracy
4	J	Rotational quantum number
5	Γ_{tot}	Total symmetry in $\mathcal{C}_{3\mathrm{v}}$(M)
6	K	Rotational quantum number, projection of J onto the z-axis
7	Γ_{rot}	Symmetry of the rotational contribution in $\mathcal{C}_{3\mathrm{v}}$(M)
8	L	The projection of the total vibrational angular momentum
9,10,11,12	$n_1 - n_4$	Normal mode vibrational quantum numbers
13,14	L_3, L_4	projections of the angular momenta corresponding to n_3 and n_4
15	Γ_{vib}	Symmetry of the vibrational contribution in $\mathcal{C}_{3\mathrm{v}}$(M)
16	$\lvert C_i\rvert^2$	Largest contribution used in the assignment
17,18,19,20,21,22	$s_1, s_2, s_3, b_1, b_2, b_3$	Local mode vibrational quantum numbers

Table 5.3: Extract from the SAlTY Transition file.

F	I	$\mathrm{A_{IF}}$ / s^{-1}
4220641	4736989	9.0696e-04
8442759	8640461	5.3636e-05
1269889	1056999	5.5676e-04
4631869	4737012	2.3014e-04
4632512	4737066	9.6883e-04
614599	820125	1.0712e-03
3549641	3825894	9.3653e-04
8085571	7937418	5.3630e-07
2304706	2606502	2.4236e-03
3829402	3545923	1.0250e-04
1750096	1497115	2.3840e-04
823228	612463	4.8085e-07
7589341	7582878	8.6990e-04
507260	612492	5.5009e-04
6611560	6605474	2.2927e-04
2306005	2300595	1.8031e-04
5869339	5986016	1.5085e-04
870236	1057299	4.1518e-04

I: Upper state counting number; F: Lower state counting number; A_{IF}: Einstein A coefficient in s^{-1}.

5.2 Background to the Calculation

The SAlTY calculations used an increased refinement of the *ab initio* (CCSD(T)/aug-cc-pV(Q+d)Z) potential energy surface (PES) [127] used for the SYT line list. This refinement was necessary because the PES is an 'effective' surface, so it is inextricably connected to the size of the basis set. To achieve the accuracy and degree of completeness required of the SAlTY line list, the present calculations used an increased ro-vibrational basis set, so a refinement of the surface was necessary. It was done by performing a least square fit to available experimental ro-vibrational energies of PH_3 with the rotational quantum numbers $J = 0, 1, 2, 3, 4$, mostly taken from the HITRAN database, as described in chapter 3.3.5.

The PES parameters used here are given as Supplementary Material to the associated SAlTY paper in the form of a Fortran 95 program. It should be noted that this is also an 'effective' PES and guarantees to give accurate results only in conjunction with the same method and basis set used to produce it with.

As with the SYT PH_3 line list, the variational rotation-vibration program suite TROVE [193] was employed for all nuclear motion calculations for SAlTY. The application of variational methods to polyatomic molecules requires increasingly large Hamiltonian matrices to be diagonalised, which is very computationally demanding and until recently prohibitively so. Is is only with the recent power and parallelism of modern computers that it is possible to use variational methods for the production of accurate spectra.

To accommodate the higher demands of the present line list, a larger basis set was used, with a corresponding higher polyad number. The polyad number, P, controls the size of the basis sets at all contraction steps using the polyad-truncation scheme [128], as described in chapter 3.3.3.

TROVE [193] solves the Schrödinger equation for the rotation-vibration motion of nuclei to obtain eigenvalues (ro-vibrational energies) and eigenfunctions (nuclear motion wavefunctions). The latter are necessary for ro-vibrational averaging of the dipole moment of the rotating molecule and thus to compute the transitional probabilities, usually expressed in terms of the Einstein coefficients or line strengths following the description by [192]. Further details of the TROVE computational procedure are given by Yurchenko et al. [193].

The SAlTY basis set contains two contributions, (i) all basis functions with the primitive quantum numbers satisfying $P \leq 16$ and (ii) stretching functions ranging up to $P = 20$ but with some high P-polyad ($P \geq 17$) stretching contributions that couple all three stretching

modes removed. Increasing the polyad number is computationally costly, and as such the increased coverage of the stretching excitations only was motivated by assumption that the stretching excitations produce the strongest transitions. The larger basis set guarantees a better convergence for the present calculations, but requires the PES to be refined to the new basis set, as discussed previously. The $J = 0$ Hamiltonian matrices were then constructed, and the empirical basis set correction scheme (EBSC)[185] described in chapter 3.3.6 was applied. Here, band centre values from the ro-vibrational calculations are replaced with extremely accurate corresponding experimental values, or deviated towards these values. This is performed iteratively, until the entire band is considered to be optimal. At the end of this correction, the rms deviation for the bands whose centre had been replaced went from an rms deviation of 0.02 cm^{-1} in SYT to 0.012 cm^{-1} in SAlTY. This improvement may reflect the better use of EBSC, but overwhelmingly is due to the use of the enhanced PES. Table 5.4 gives a band-by-band summary of this improvement. The $2\nu_4$ and $\nu_2 + \nu_4$ bands were particularly responsive to this process.

Table 5.4: Observed bands centres, from HITRAN [139], and standard deviation, σ, with which the TROVE calculations reproduce the terms within each band. The first three columns give integrated band intensity for each band calculated for SYT, SAlTY and HITRAN.

Band	Band Centre	σ (cm^{-1})		Band Intensity (cm / molecule)		
	(cm^{-1})	SYT	SAlTY	HITRAN	SYT	SAlTY
ν_2	992.135	0.020	0.015	3.087×10^{-18}	3.264×10^{-18}	3.261×10^{-18}
ν_4^1	1118.307	0.005	0.003	3.149×10^{-18}	3.436×10^{-18}	3.436×10^{-18}
$2\nu_2$	1972.571	0.007	0.004	9.390×10^{-21}	1.414×10^{-20}	1.468×10^{-20}
$\nu_2 + \nu_4^1$	2108.152	0.034	0.013	9.504×10^{-20}	1.484×10^{-19}	1.487×10^{-19}
$2\nu_4^0$	2226.835	0.010	0.008	4.689×10^{-19}	7.684×10^{-19}	7.590×10^{-19}
$2\nu_4^2$	2234.915	0.014	0.005	combined above	combined above	combined above
ν_1	2321.121	0.012	0.008	4.926×10^{-18}	6.091×10^{-18}	6.127×10^{-18}
ν_3^1	2326.867	0.013	0.008	1.454×10^{-17}	1.650×10^{-17}	1.648×10^{-17}
$3\nu_2$	2940.767	0.046	0.032	6.725×10^{-21}	2.127×10^{-20}	2.129×10^{-20}
$\nu_2 + 2\nu_4^0$	3214.936	0.024	0.017	3.612×10^{-21}	7.478×10^{-21}	6.996×10^{-21}
Overall		0.020	0.012			

Table 5.4 also compares the integrated band intensity computed for the SYT and SAlTY line lists with those obtained from HITRAN 2012 [139]. In each case, the band intensity was computed by explicit summation of the intensities at 296 K of all lines within a band. The two $2\nu_4$ bands were considered together since it is difficult to disentangle their transitions. The band intensities for the two computed lists are very similar. They are both in reasonable agreement with, but somewhat larger than, HITRAN; given the greater completeness of the computed linelists, it is to be expected that this method would yield somewhat larger band intensities.

For the rotational spectra from CDMS the integrated intensity is 1.214×10^{-18} cm/molecule which is very similar to SAlTY's 1.218×10^{-18}. One further test on the intensities was performed. Given that the intensity of weak lines can be highly sensitive to the choice of wavefunction [162], and hence the underlying PES, an explicit comparison was made between SYT and SAlTY of the intensities of the weak, forbidden lines with $\Delta K = 3$. These were found to be very little changed for the individual transitions inspected, with the integrated intensity for all these transitions going from 5.53×10^{-19} cm / molecule in SYT to 5.60×10^{-19} cm / molecule in SAlTY. In practice variational procedures, such as those use here, have long been used to get reliable predicted intensities for such transitions [108, 93].

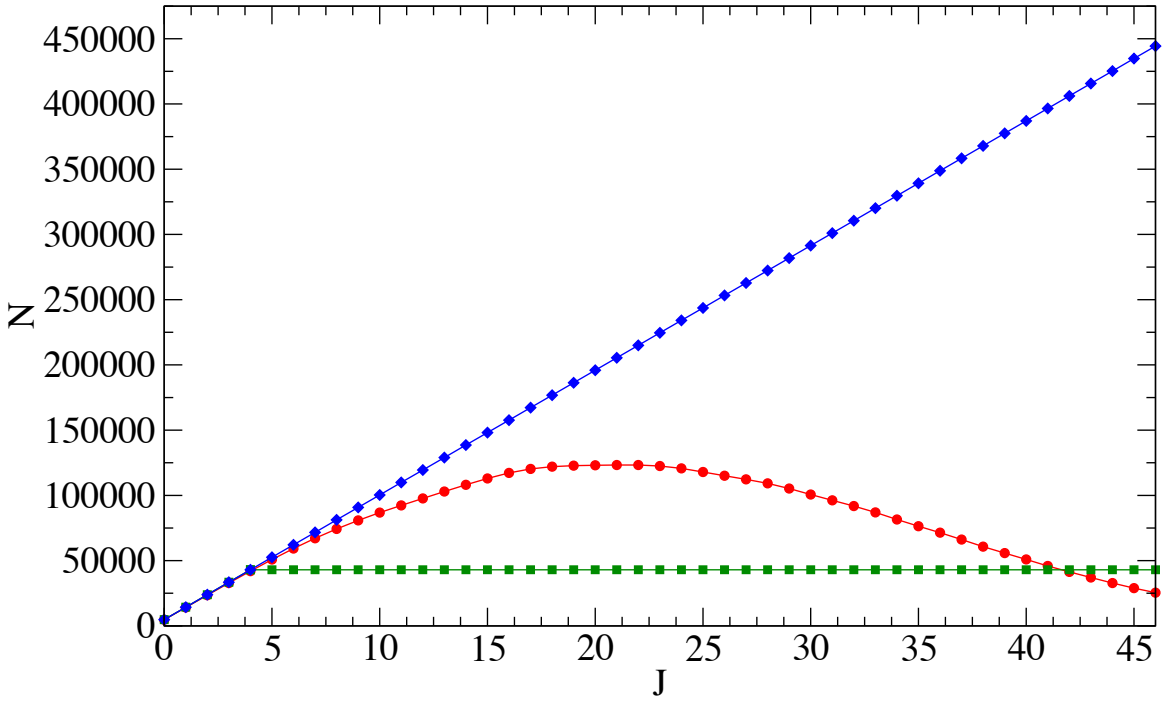
5.2.1 Energy Eigenvalues

As well as employing a larger basis set than before, this line list improves on SYT by using (i) a larger energy level range of $E_{max} \leq 18\,000$ cm^{-1}(instead of $\leq 12\,000$ cm^{-1}), (ii) wider frequency range of 0 – 10 000 cm^{-1} (instead of 0 – 8 000 cm^{-1}) and (iii) rotational excitations considered up to $J_{max} = 46$ (instead of 34). This expansion is necessary to guarantee accuracy and completeness at high temperatures, but it makes the calculation of the line list much more computationally demanding.

The diagonalisation of the matrices involved in the variational procedure corresponding to high values of J is the most computationally demanding part of a line list calculation, as it requires substantial memory, long runs and requires MPI to diagonalise efficiently. A variety of strategies were used to make the diagonalization of all the SAlTY matrices possible, and are described below.

The dimension of the matrices of the symmetries scales approximately as 1:1:2, for A_1, A_2 and E, respectively. As can be seen from Fig. 5.1, the size of the ro-vibrational matrices scales linearily with J, but the number of non-zero elements and the number of eigenvalues under the energy threshold for the line list (8 000 cm^{-1}) do not. The size of the matrices grows roughly with $N^{\Gamma}_{J=0} \times (2J+1)$, where $N^{\Gamma}_{J=0}$ is the dimension of the matrix block for $J = 0$ with symmetry Γ. For the symmetry with the biggest matrices, $\Gamma = E$, $N^{E}_{J=0}$ = 4778. The largest matrix to be diagonalised for SAlTY ($J = 46$, E symmetry) has dimensions of 444 726.

Figure 5.1: Dimensions of the E-symmetry matrices with J (blue diamonds), the corresponding number of eigenvalues below 18 000 cm^{-1} (red circles) and number of non-zero elements on each row (green squares).



The initial (low J) matrices were sufficiently small to employ the standard LAPACK eigensolver DSYEV [9] to solve the full eigenvalue problem, and this was used for all $J \leq 16$ and 9 (A and E symmetries, respectively). The OpenMP PLASMA library [81] was used for matrices with dimensions between 100 000 and 200 000. As the matrices increased in size, memory constraints required the MPI version PDSYEVD of this eigensolver to be used on large shared memory systems (COSMOS) in order to cope with the size of these matrices.

Using these three eigensolvers, all eigenvalues and eigenvectors were computed for $J \leq$

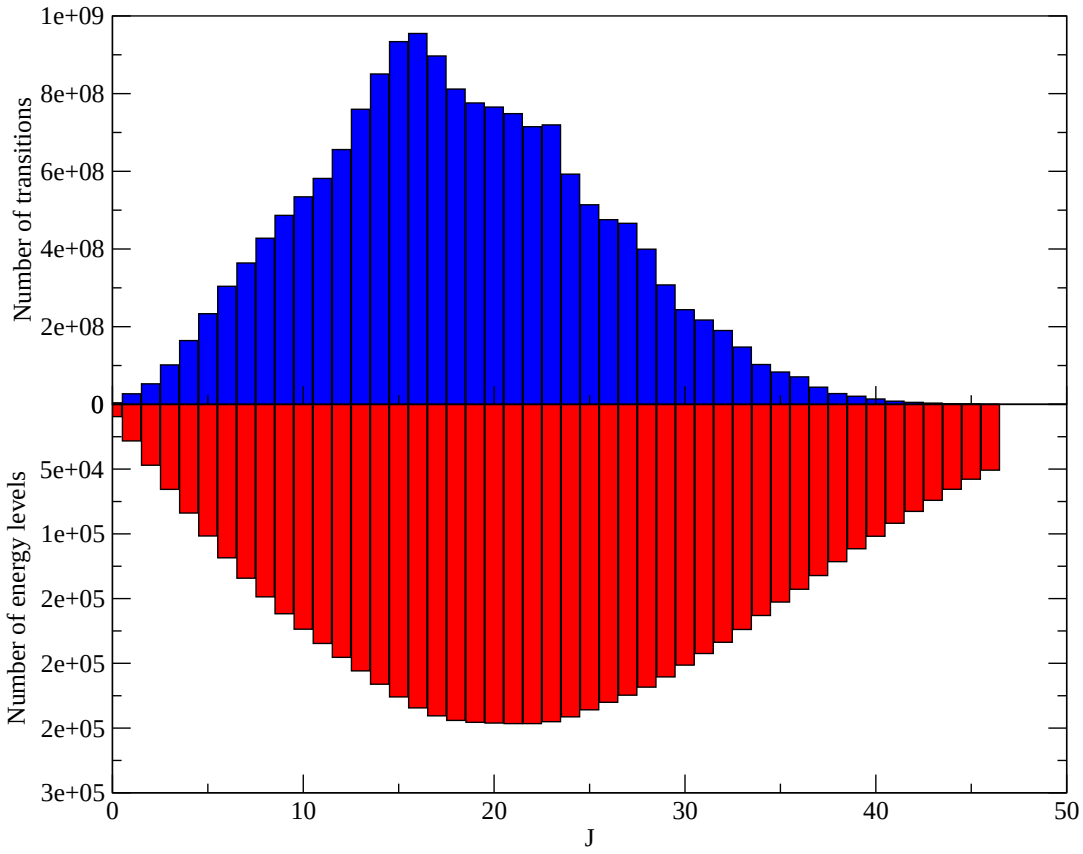
46, using the energy threshold of 18 000 cm^{-1}, totalling 7 480 690 energy levels.

5.2.2 Transition Strength Calculations

A high quality dipole moment surface (DMS) is a prerequisite for correctly computing the transition moments used to generate the Einstein A coefficients and in turn accurate line intensities; Tennyson [162] gives a discussion of the issues involved in this. The DMS used here is completely theoretical, an approach that has been shown to yield results competitive with experiment [98], and has already been used in the SYT line list where it provided intensities in good agreement with experiment. It is a six-dimensional *ab initio* (CCSD(T)/aug-cc-pVTZ) electric dipole moment surface, calculated on a grid of 10 080 molecular geometries, and it is described in further detail elsewhere [187]. To reduce the size of the computation, only eigen-coefficients of the upper states with a magnitude larger than 10^{-14} were selected. After this work was completed, Nikitin et al [119] published a detailed analysis of *ab initio* methods for calculating the phosphine DMS. In general the agreement between their predictions and ours is very good.

Despite matrix diagonalization utilising the most computational resources, it is the computation of transition intensities that actually dominates the total computer time. For example, there are approximately 500 million transitions between E-symmetry energy levels in $J = 22$ and $J = 23$. These transition eigenvectors are extremely dense, with a basis size is of the order of 700000; the estimated time required to compute the $22 \leftrightarrow 23$ transitions is in the order of 934 hours or over 1 month. This is longer than the wall-time limit in most computational clusters. However, the calculation of intensities has the advantage that it is easily paralellised and spread over many processors, and indeed computer systems. Nonetheless, the sheer number of transition calculations between J pairs can still be debilitating in terms of computer and real time. As shown in Fig. 5.2, the number of transitions between energy levels peaks with $J = 15$ - 18, so this was the region that corresponded to the most expensive part of the intensity calculations, accounting for about 17% of the total transitions. The number of transitions does not correlate smoothly with J pairs due to the different way A and E symmetries are affected by J, particularly for values of K that are multiples of 3.

Figure 5.2: Number of energy levels(red, below) in each rotational quantum number, J, and transitions(blue, above) between J and J,$J+1$ in the SAlTY line list, summed over all symmetries.



Employing CPUs available to the ExoMol group at the HPC centres DiRAC@Darwin and DiRAC@COSMOS, TROVE was used to calculate the intensities across almost all symmetries and J values. However this proved slow and computationally expensive, so for some of the particularly computationally demanding sections, a newly-developed graphical processor unit (GPU) implementation of the TROVE program was employed. This allowed us to use multiple GPUs to compute multiple initial states simultaneously, effectively reducing computing time by over an order of magnitude. Additionally, further parallelisation was introduced by splitting the calculation of the transitions, not just between J pairs, but by lower energy states, since these are calculated independently. However, this increases the risk for the creation of duplicate transitions, which can be onerous to detect and remove. This update to the intensity calculation process in TROVE is called GAIN; the algorithm it employs is described by Al-Refaie et al [3]. GAIN was used to compute the absolute intensities for most

A symmetry transitions with $J \geq 20$, on the Emerald CfI cluster. It was also used to compute the intensities for a few of the E-symmetry J pairs; for example, the previously mentioned transitions for J 22 $\leftrightarrow$ 23 were completed in 56 hours using an 8 GPU node.

At the end of the SAlTY calculations, a total of 16.8 billion phosphine transitions were simulated.

5.3 Linelist Validation and Temperature Dependence

Line lists do not specify a temperature, since the Einstein coefficients for the transitions are independent of temperature. SAlTY can only be considered a 'hot' line list because its energy and rotational excitation thresholds ensure that all those states that are significantly populated up to $T = T_{\max}$ are used to produce the catalogue of transitions.

As discussed in chapter 2.1, to model the spectrum of phosphine at different temperatures, the line intensities are calculated using:

$$I(f \leftarrow i) = \frac{A_{if}}{8\pi c} g_{\text{ns}} (2J_f + 1) \frac{\exp\left(-\frac{E_i}{kT}\right)}{Q\ \tilde{\nu}_{if}^2} \left[1 - \exp\left(\frac{-hc\tilde{\nu}_{if}}{kT}\right)\right], \tag{5.2}$$

where k is the Boltzmann constant, T the absolute temperature, g_{ns} is the nuclear spin statistical weight factor and Q is the partition function. Chapter 4 describes in depth the temperature dependent partition function of phosphine. The work associated with it produced accurate and fully converged values for the partition function for temperatures below 3000 K by the explicit summation of $\approx$ 105 million theoretical rotation-vibration energy levels of the molecule. The convergence dependence on temperature arises due to the growing contribution higher states make towards the value of the partition function as temperature increases.

The $T_{\max}$ for which molecular data is complete to a satisfactory degree can be found by computing the temperature-dependent partition function using only the energy levels under the lower energy threshold considered in the line list (here 8 000 cm^{-1}), and then comparing its value to that of the complete partition function described in chapter 4. Completeness of the line list with temperature is measured by considering the ratio $Q_{\text{limit}}/Q_{\text{total}}$, where Q_{total} is the converged partition function value calculated by explicitly summing over all energy levels and Q_{limit} is the partition function calculated using only those levels with energies up to SAlTY's threshold of 8 000 cm^{-1}. Figure 5.3 shows SAlTY's completeness with increasing temperature. SAlTY is over 90% complete for temperatures below 1500 K, but quickly becomes depleted at higher temperatures. It is possible to use SAlTY to model temperatures

over 1500 K by estimating the proportion of opacity missing, using the percentual loss of completeness [114] in Figure 5.3, but it is recommended that 1500 K is taken as a soft limit to the applicability of the SAlTY line list.

Figure 5.3: Ratio of effective partition function used in SAlTY, Q_{8000}, to the converged value, Q_{total} (previously calculated, see Chapter 4). This ratio gives a measure of completeness of SAlTY as a function of temperature.

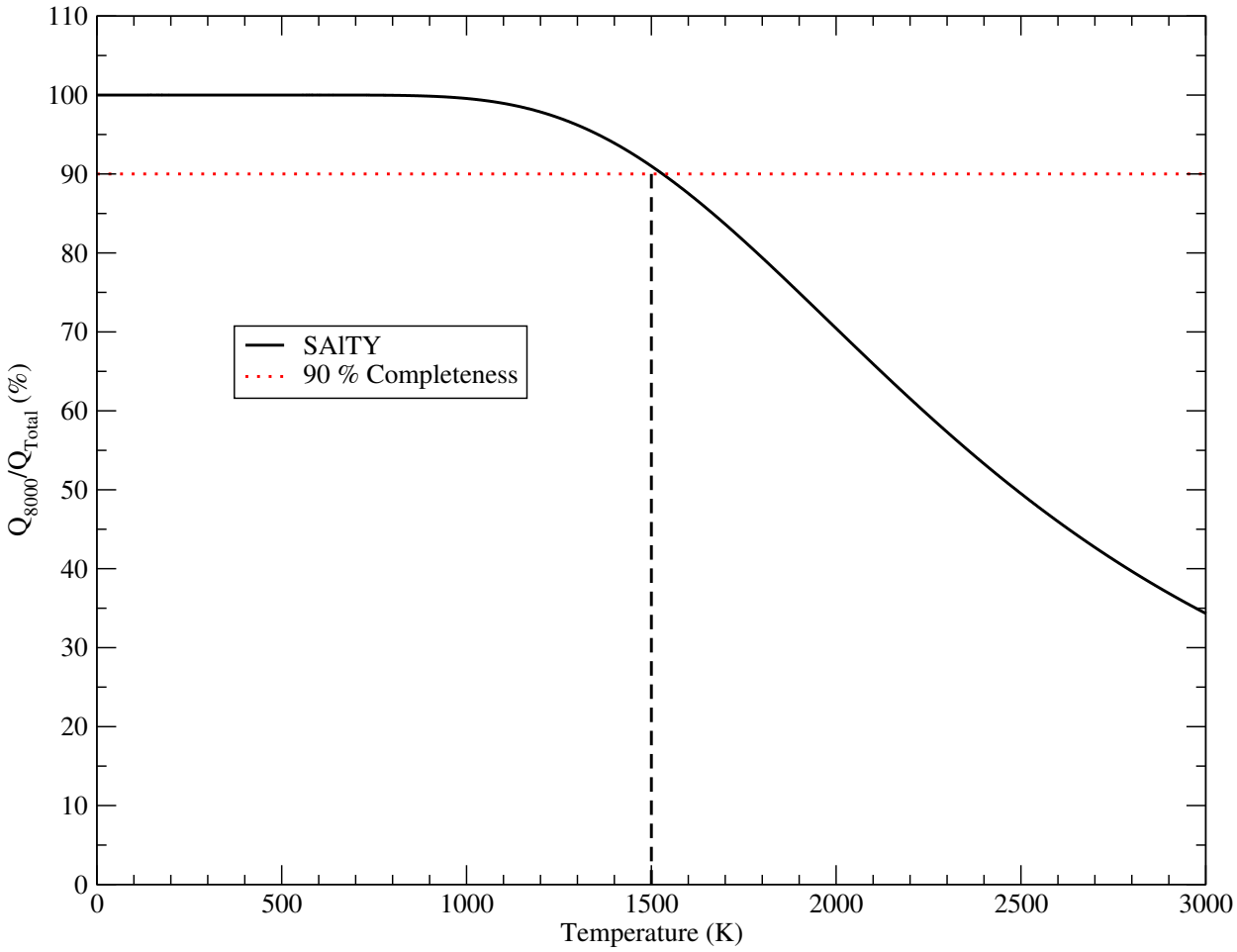


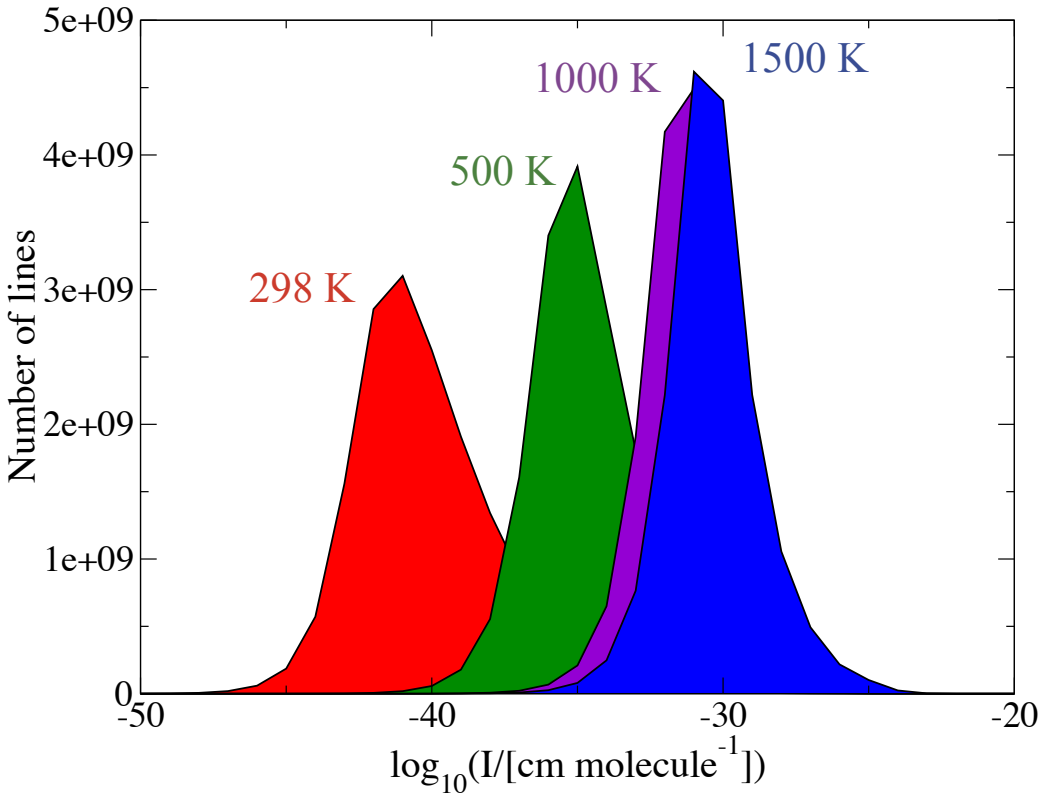
Table 5.5 provides a summary of SAlTY's completeness with temperature, where it can be seen that for temperatures as high as 2500 K, SAlTY will only be 50% complete.

Table 5.5: Maximum temperatures for which the SAlTY line list is percentually complete.

$T_{\max}$ (K)	Completeness %
1014	100
1146	99
1797	80
1500	91
2000	70
2500	50

Fig. 5.4 demonstrates how the density of lines per absorption intensity $I_{if} = A \times 10^x$ unit changes with temperature, covering the whole wavenumber range 0–10 000 cm^{-1}.

Figure 5.4: Number of intense lines as a function of intensity for different temperatures. The x-axis gives the log of the intensity in cm/molecule, while the y-axis represents the number of transitions per each 10^x cm/molecule bin.



As the temperature rises, the number of intense lines increases but the range of intensities in the spectrum becomes narrower. The Gaussian-like intensity distributions peak at I =10^{-41}, $= 10^{-35}$, $= 10^{-31}$, and $= 10^{-31}$ for T = 298, 500, 1000, and 1500 K, respectively. This is different from the intensity distribution found for CH_4 by Yurchenko *et al* [190], where the proportion of strong lines was found to be much larger.

5.3.1 Cross-sections

Using the SAlTY line list, multiple cross-sections were simulated for PH_3. Figures 5.5 and 5.6 illustrate the temperature dependence of phosphine for T = 300, 500, 1000, 1500 and 2000 K. It can be seen that the spectra at higher temperatures becomes smoother, with less features and a loss of sharp Q-branches.

Figure 5.5: Overview of the full SAlTY spectrum of PH_3 for T = 300, 500, 1000 and 1500 K, absorption cross-sections (cm^2/molecule) with HWHM = 0.5 cm^{-1}. Looking at the minimum of the spectra, the cross-sections are ordered in increasing temperature.

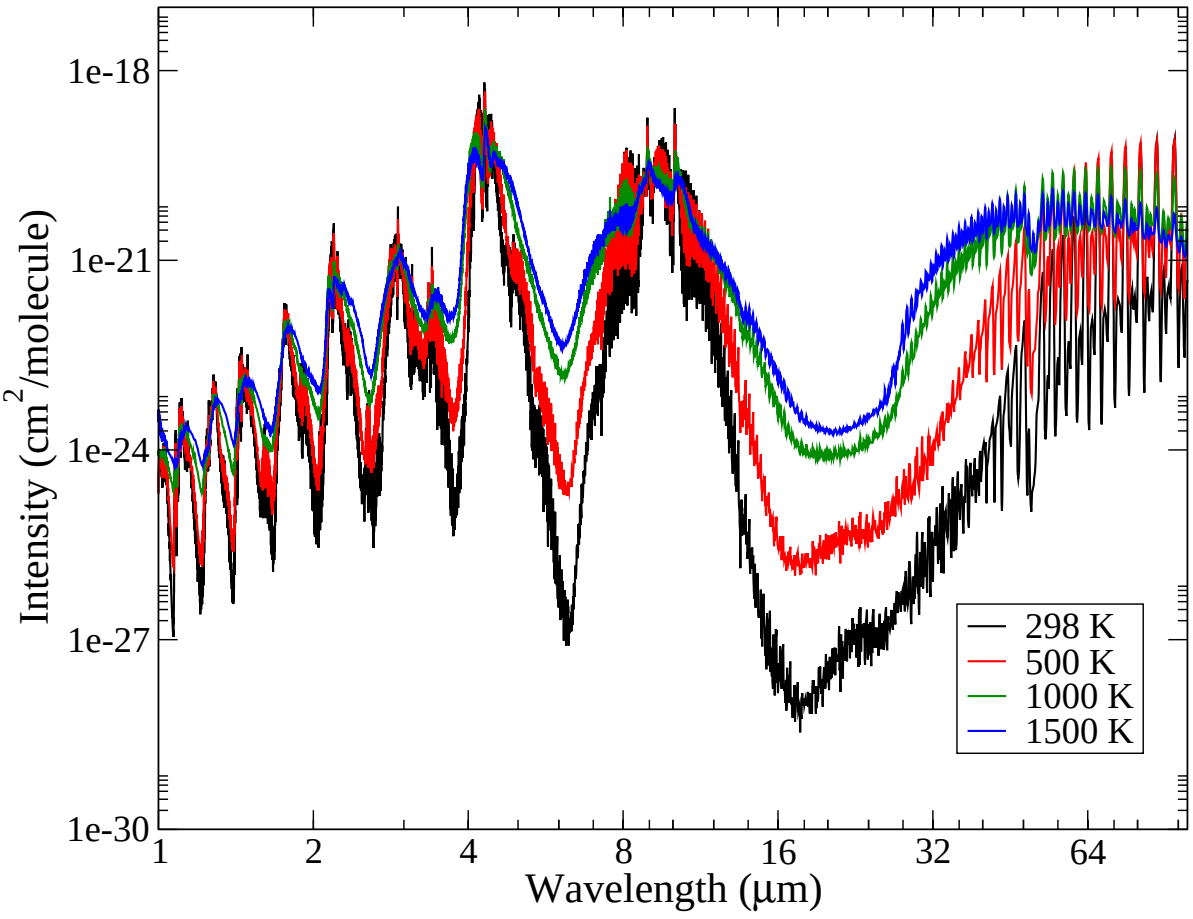


Figure 5.6: SAlTY absorption spectra of PH_3 for T = 300, 500, 1000, 1500 and 2000 K, convoluted with a Gaussian profile, HWHM = 2 cm^{-1}, for the 30(**a**), 10(**b**), 5(**c**) and 3(**d**) μm regions.

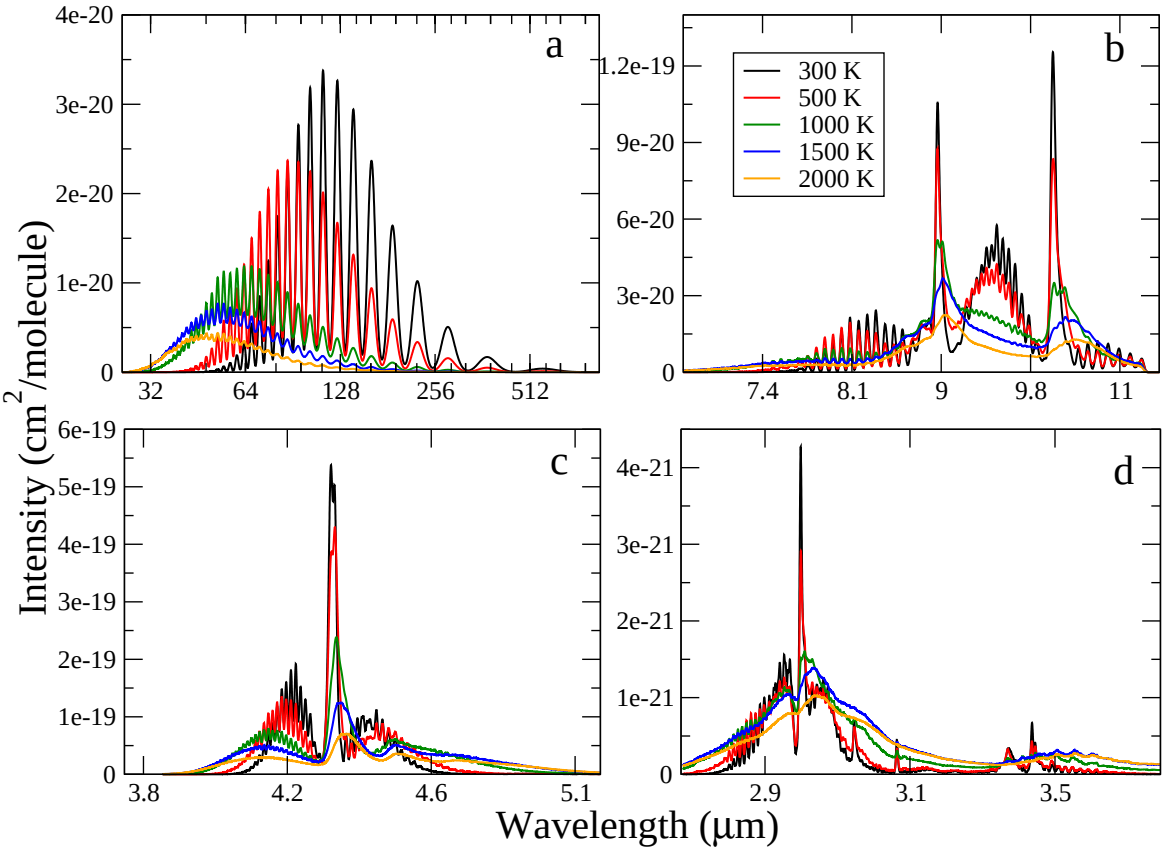


Figure 5.6, part **c**, covers the wavelength region 3.8 – 5.1 μm which, as mentioned before, is of particular importance for the spectral characterisation of brown dwarfs and gas giants.

5.3.2 Comparison with Experiment

Chapter 3 offers a detailed comparison of SYT to the existing experimental data, where it was found that it replicates very well the observed phosphine spectra at room temperature, with a maximum rms deviation from CDMS [112] of 0.076 cm^{-1} for the rotational spectrum and of 0.23 cm^{-1} from HITRAN 2008 [140]. The most recent CDMS update [112] contains a very comprehensive description of the pure rotational band of phosphine. Figure 5.7 demonstrates that SAlTY is in excellent agreement with the CDMS data. The transitions in the bottom left box shows a deviation of approximately 0.01 cm^{-1}, while the right most box, at higher wavenumbers, shows a deviation of approximately 0.04 cm^{-1}.

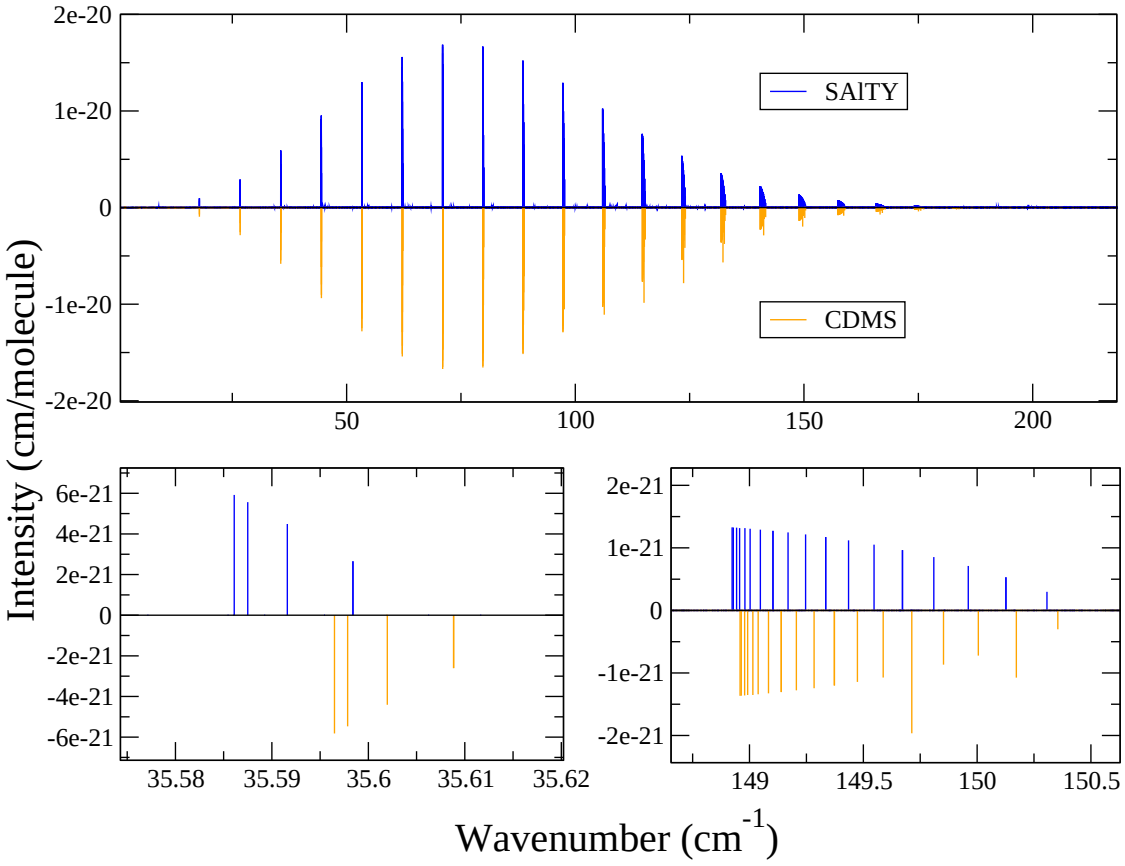
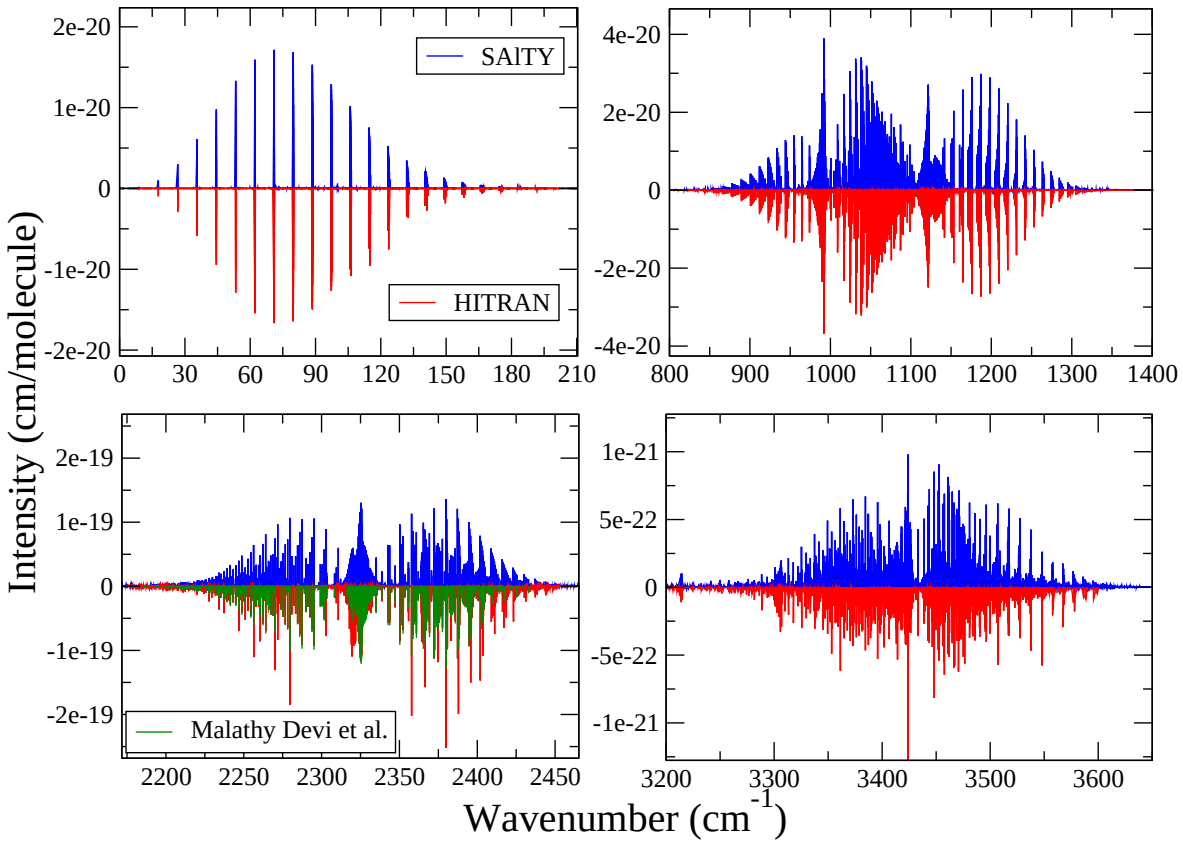


Figure 5.7: Comparison of the SAlTY line list with the most recent phosphine data from CDMS at room temperature [112].

As can be seen from figure 5.8, the SAlTY line list is also in excellent agreement with the most recent HITRAN database [139] (and incorporated CDMS data), and it is expected that SAlTY will have slightly lower rms deviations from experiment at this temperature than SYT.

Figure 5.8: Comparison of the SAlTY line list with the phosphine data from HITRAN [139] at room temperature; also shown in green are the recent results of Malathy Devi et al [103].



The apparent intensity disagreements with HITRAN at the P=2 region(1950-2450 cm^{-1}) occur because of an overestimation of the intensity of some of the degenerate $A_1 \leftrightarrow A_2$ transitions which are in such close proximity that they have been perceived as one doubly strong transition in the experimental data. This has been partially resolved by Malathy Devi *et al.* [103], who corrected these intensities for the region 1950 to 2450 cm^{-1}. The bottom left box in figure 5.8 includes this data, and it can be seen SAlTY is in much better agreement with it than with the HITRAN data. We recommend that the Malathy Devi *et al.* data should be included in the next release of HITRAN.

Other than a few exceptions relating to intensity measurements, the data in HITRAN are generally very accurate for room temperature simulations. However, if the current HITRAN database is used to simulate high temperature spectra, the differences become much more striking, as can be seen in Fig. 5.9. It is worth noting that HITRAN do not recommend that their phosphine data be used to compute spectra at high temperatures, and this comparison is only included for the benefit of those doing so against HITRAN's recommendation.

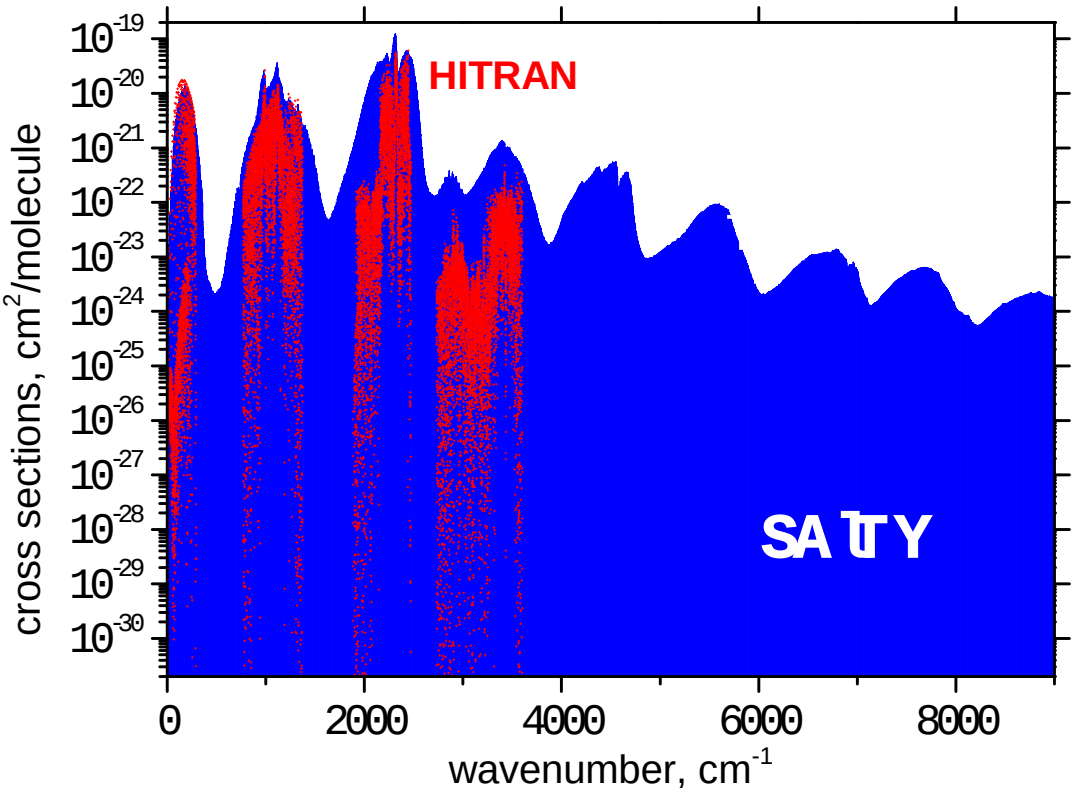


Figure 5.9: Comparison of the SAlTY line list with the phosphine data from HITRAN [139], extrapolated to 1500 K.

The only available data above 300 K comes from the Pacific Northwest National Laboratory (PNNL) who provide cross-sections for PH_3 up to temperatures of 323 K [150]. Comparisons with SAlTY for $T = 50$ ^{o}C, or 323 K, can be seen in Fig. 5.10. The PNNL data is approximately 8% weaker than SAlTY's, but both are otherwise in good agreement.

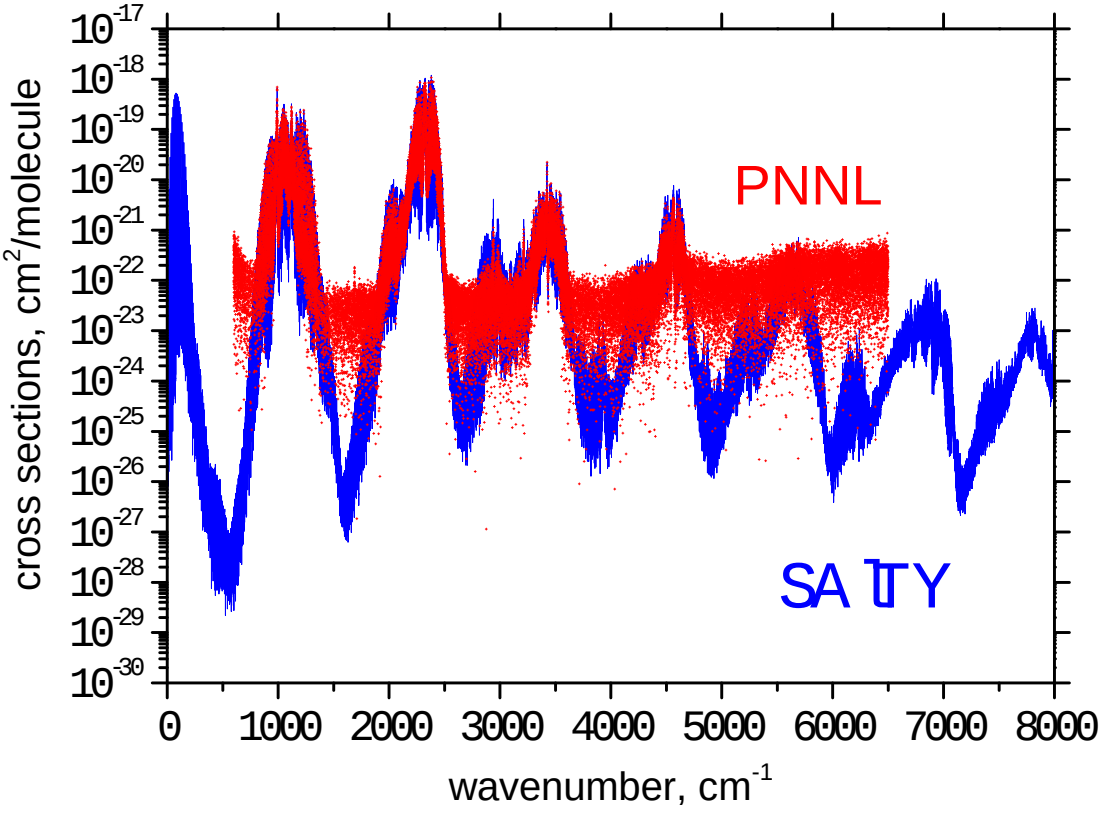


Figure 5.10: Comparison of the SAlTY absorption cross-sections for PH_3 with PNNL and HITRAN at $T = 50$ C, HWHM = 0.076 cm^{-1}.

5.4 Hot Line List Limitations

SAlTY improves on the previous room temperature phosphine line list SYT in terms of the size of the basis set, corresponding refined potential energy surface and increased spectral range. Its only limitations are: upper states with energies above 18 000 cm^{-1} are excluded due to their negligible population even at very high temperatures, and there is an effective short-end wavelength cut-off of 1 μm, which is an unimportant region for PH_3.

SAlTY contains 16.8 billion transitions between 7.5 million energy levels; this number of features is approximately five orders of magnitude larger than any existing experimental dataset for phosphine. SAlTY is suitable for accurately simulating spectra up to temperatures of 1500 K, but can be used approximately for temperatures above this.

Given that SAlTY has a larger density and coverage than any other PH_3 line list(including SYT), it is recommended that SAlTY is used to simulate all phosphine cross-sections for atmospheric spectra, even at low temperatures.

Chapter 6

Discussion and Conclusion

Phosphine has exciting potential as a both a marker for convection in gas planets and cool stars, and as a biomarker in terrestrial planets, so its existence outside the Solar System is worth investigating. To simulate the atmospheres of these environments, computational models require detailed molecular line lists; before the work presented in this thesis, no phosphine line list existed which was sufficiently complete and accurate to characterise the atmospheres of cool stars and exoplanets. The creation of SAlTY together with the development of the associated thermodynamic data allows phosphine to be modelled and detected in an unprecedented variety of environments and temperatures.

The introduction of the SAlTY line list in atmospheric models has now begun and is considered in the following section. Additionally, there are two current and future projects arising from the simulation of phosphine. The first is the Phantom Symmetries project, which is a process of exploiting the symmetry properties of molecules to simplify the creation of low resolution spectra, using phosphine as preliminary example. The second is the exploration of the tunnelling motion of phosphine, which is of particular interest to the field of molecular structure, due to phosphines extremely small splitting, which is as of yet undetected. These two applications are discussed in Sections 6.2 and 6.3.

6.1 Applications in Astronomy

Phosphine is expected to play a crucial role in both LTE (local thermal equilibrium) and non-LTE environments outside the solar system (see section 1.3). It has been detected in the circumstellar envelope of IRC +10216 and is a convective marker in the atmospheres of Jupiter and Saturn, but the list of astrophysical locations of phosphine has not expanded significantly in the past decades, partially due to the relative lack of adequate molecular data.

Phosphine could be an important molecule outside the Solar System for a few reasons. First, as an integral part of the phosphorous cycle in the biospheres, hydrosphere, and atmosphere of the Earth [42, 196, 34], it is likely that it will also play a relevant role in these processes in terrestrial exoplanets. Second, phosphine is a phosphorous species particularly important in desequilibrium chemistry, as its destruction mechanism and quenching lead to major abundance variations within a non-LTE environment. This can occur both through vertical fixing and horizontal transport, the latter of which is relevant for transit observations of terminator regions [111]. This role phosphine plays as a marker for convective dynamics makes it also useful for probing the deeper layers of the atmospheres [159]. In an analogous circumstance to the Jovian and Saturn atmospheres, phosphine could play an important role in the convective regions of brown dwarfs and giant hot planets where, for temperatures greater than 1300 K, the opacity due to H_2-H_2 is significantly increased, making it hard to probe atmospheric layers in depth [123].

Accurate detections and predictions of phosphine in the environments mentioned above will require comprehensive molecular data, and SAlTY represents a considerable improvement on the previous experimental phosphine line lists. This is most evident when a cross-sectional spectra is created and the collective power of the millions of weak lines becomes apparent. Incorporating such an extensive catalogue of transitions in a computational model requires adequate cross-sections and associated line functions, which are discussed below.

6.1.1 Phosphine Line Widths

The transitions in the SAlTY line list have zero line widths, as the theoretical lines have not yet been broadened. When using SAlTY to create atmospheric models that can be compared to observations, the lines have to be widened to better simulate the studied environment's physical characteristics. This is further detailed in section 2.1.

Even an isolated spectral line will experience Doppler broadening and some self-broadening, and no real spectra is completely isolated. It is therefore important to fit an appropriate line function that will account for the change in spectral shape caused by the effect of local pressure and temperature. These line functions are difficult to create and currently, only zero-pressure temperature-dependent molecular absorption cross sections exist for SAlTY, with the Doppler shift effect accounted for using a Gaussian function to shape the transition lines [71]. These are freely available on the ExoMol website (*www.exomol.com*), with a variable bin size.

Pressure broadening of a line is heavily dependent on the transition's J numbers, but the K value and the vibrational quantum numbers cause no significant variation within each J pair, except when K approaches J [88]. At K=J, PH_3 rotates about its principal symmetry axis, which leads to a longer lifespan of the states corresponding to this arrangement. For smaller values of K, the rotation axis deviates from the principal symmetry axis, giving these states a short duration and consequently larger broadening.

As part of future work, phosphine line-broadening due to H_2, He, Ar, and self-broadening will be explored. In the meantime, there are some experimental measurements of broadening coefficients for a few strong phosphine lines, described below.

There are measurements of H_2, He, Ar and self broadening coefficients for a few strong lines in the ν_2 and ν_4 bands in the the polyad 1 region (section **3.4.5.1**) [143, 144, 21]. These measurements have also been used to calculate the temperature dependence of collisional broadening coefficients for H_2-PH_3, but all measurements and calculations have been done with low temperatures (T$\leq$300 K) in mind. H_2 line mixing coefficients were measured for 32 strong lines belonging to the ν_2 and ν_4 bands in the range 1016 – 1106 cm^{-1}[144]. It was found that the line mixing parameters for PH_3 are small when compared to NH_3, and have no systematic dependence on J. Additionally, the ν_2 band is more strongly affected by mixing than the ν_4 band.

Malathy Devi *et al.* has measured line self-widths, self-shifts and mixing coefficients for nearly 840 transitions in the ν_1 , ν_3, $2\nu_4$ and $\nu_2 + \nu_4$ band, covering the 2100 - 2500 cm^{-1} region [103]. They also present a summary of the existing widths, shifts and line mixing measurements for the studied PH_3 bands. Additionally, some of the transitions in the HITRAN catalogue contain spectral line shape parameters [139], but Malathy Devi *et al.* are at present the only source of pressure-induced shift coefficients for any broadening gas for phos-

phine. These data could help create a theoretical model of the pressure-broadening effect in phosphine.

Due to the absence of rate coefficients for collisional excitation of phosphine by H_2 and He, previous work has used those computed for NH_3, correcting for the near absence of inversion doubling [1]. This is a possible approach to follow but it should be noted that this introduces a large uncertainty to the calculations.

Better understanding of the coefficients that affect the line shape of PH_3 transitions, as well as their temperature dependence, becomes increasingly important for the accurate analysis of high resolution spectra. Future work will improve on the line shape description of phosphine but in the meantime SAlTY can be incorporated into models as a set of Doppler broadened cross-sections.

6.1.2 Atmospheric Modelling of Phosphine

To paraphrase Brooks and Maja [23], the problem with modelling is that is it doomed to succeed. Modern computational models of exoplanet atmospheres use the best available fundamental molecular data and sophisticated line shape functions. Often this complex modelling requires a large number of free parameters which inherently better the fit between a model and the observational data; this is not a marker for an improved model as the addition of every new parameter provides the possibility for an automatic enhancement of the fit which does not necessarily provide greater explanatory power [163]. Modellers must therefore be careful to maintain their free parameters within a credible range and perform strict statistical analyses to ensure the validity of their fit.

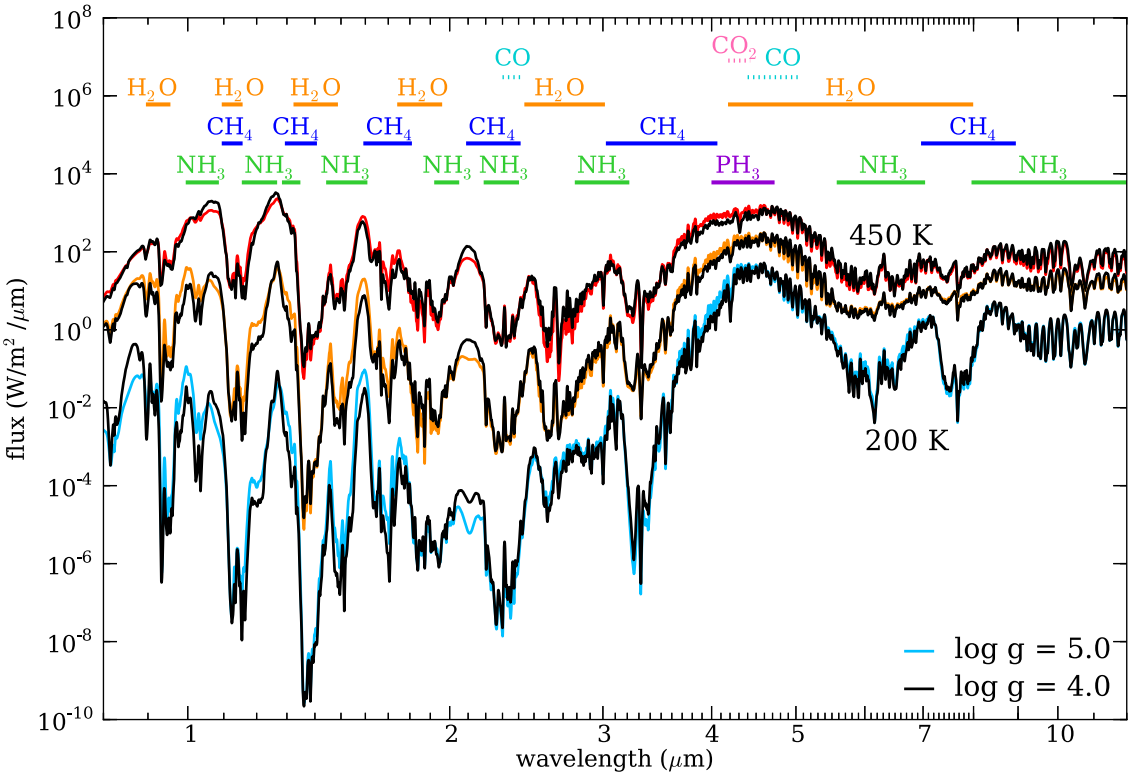
The creation of absorption cross-sections from a line list is of great importance as most radiative transfer models are not equipped to use zero-width transition line lists, nor would it be practical to input billions of lines for each molecule used in the model. Of course, there are issues creating cross-sections from line lists for use in astrophysical applications as it requires collaboration between people with very different sets of skills. The interdisciplinary nature of the ExoMol team makes it particularly well suited to work in this interface between molecular simulations and atmospheric modelling.

The detection of phosphine on an atmospheric spectra depends on whether it has strong enough features in spectral regions where the opacity from other molecules is sufficiently low. Water clouds, for example, are optically thick for objects below 375 K, but do not strongly

impact an atmospheric spectra for temperatures above this [110]. This is despite water condensing at high altitudes for any objects with temperatures below $\approx$ 400 K. Phosphine can still contribute to the overall opacity of a spectrum without being a dominant feature.

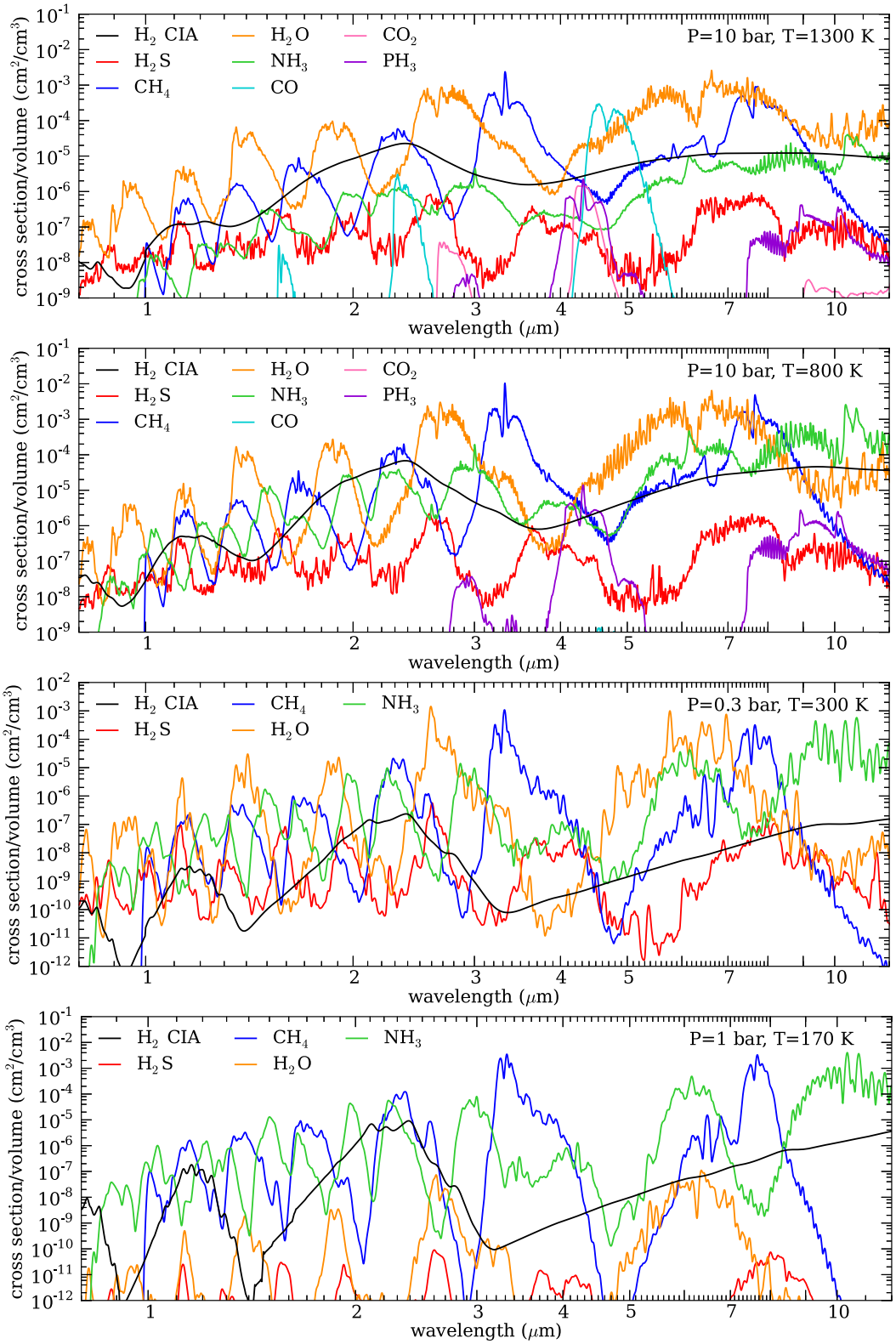
As shown in Figure 6.1, for most Y dwarfs, phosphine is predicted to be an important absorber in the infrared, with H_2O, CH_4 and NH_3 being the dominating sources of opacity elsewhere. It is worth noting that these Y dwarf models were performed without SAlTY, so the molecular data could have been insufficient to represent the true absorbance of phosphine.

Figure 6.1: Summed flux for Y dwarfs at T_{eff} = 200, 300 and 450 with log gravities 4.0 and 5.0. Phosphine is the dominant opacity source at 4.3 μm. Figure credited to Morley et al [110].



Morley *et al* [110] also modelled cross-sectional spectra for brown dwarfs of much higher temperatures where the opacity due to phosphine, although still seen as important in the infrared for T dwarfs, is likely to have been underestimated by using experimental data recorded at room temperature, rather than SAlTY. Figure 6.2 shows the spectra of the main constituents of Y and T dwarfs, presuming solar number densities for all molecules and local thermal equilibrium, where phoshine is predicted to be a dominant absorber.

Figure 6.2: Cross-sectional spectra for the opacities of H_2, H_2O, H_2S, CH_4, NH_3, PH_3, CO_2 and CO, in Y and T dwarfs. Figure credited to Morley et al [110].



SAlTY has only just began to be incorporated in atmospheric models, but its impact has already been felt by Richard Freedman, at the SETI Institute, who has begun to use it to model temperature dependent cross-sections. The spectra were simulated using a line function program provided by Linda Brown, where the lines are broadened using H_2 and He parameters from Salem *et al.* [143]; these only exist for the ν_2 and ν_4 bands. Initially, Freedman plotted a preliminary phosphine spectra using the SAlTY line list and an intensity cut-off to dismiss extremely weak lines. As can be seen from Figure 6.3, an artificial, but significant loss of opacity is visible at high wavenumbers for temperatures as low as 750 K. This is a completeness issue, caused by the absence of the cumulative effect of millions of weak lines that were dismissed by the intensity cut-off. When the intensity cut-off was lowered to 10^{-30} cm^2/molecule, the opacity loss disappeared. Figure 6.4 shows the resulting phosphine spectra at temperatures between 300 and 2000 K, with the previously excluded transitions clearly improving the spectra at high frequencies.

Figure 6.3: Phosphine spectra with He/H_2 broadening at 1 bar pressure, using only the strongest SAlTY transitions, for temperatures of 300 K, 500 K and 750 K. Figure provided through private correspondence.

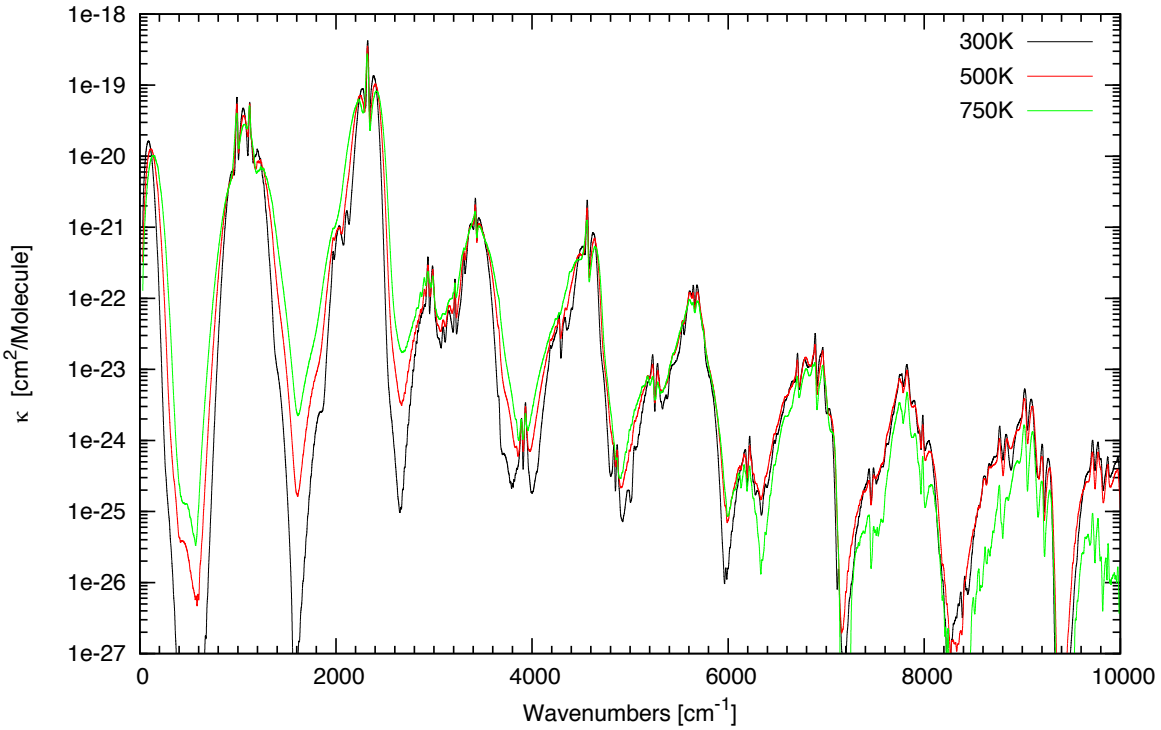
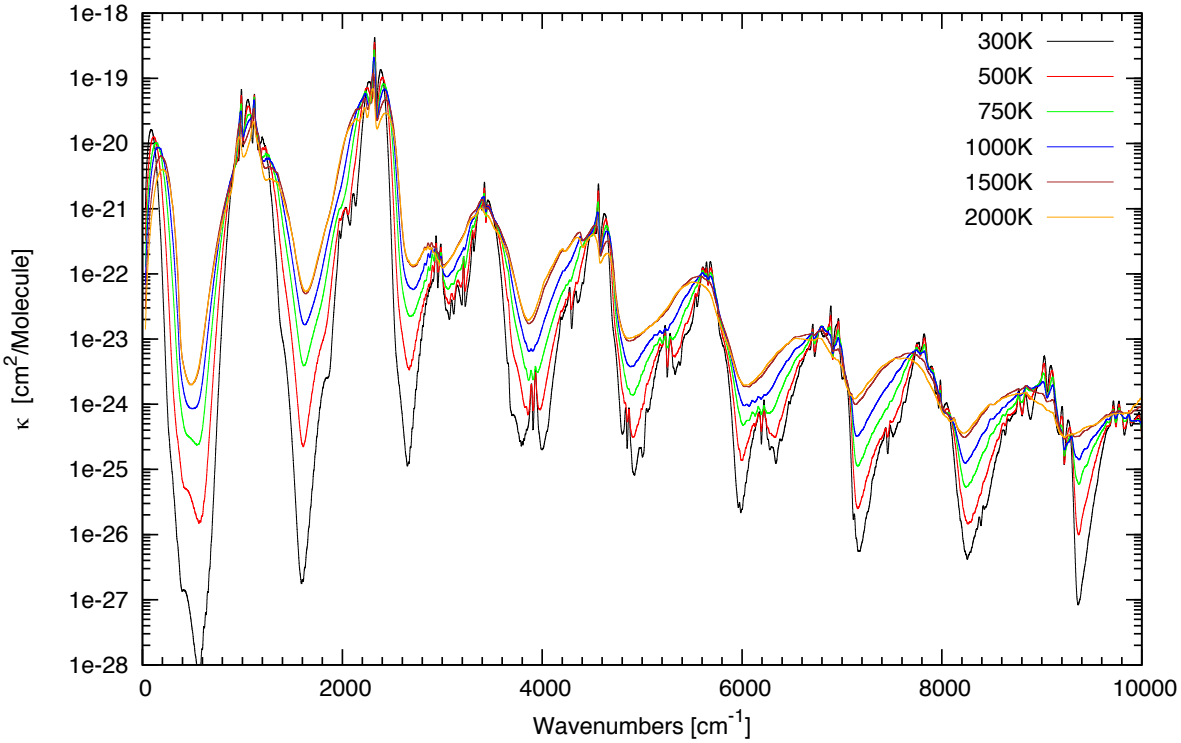


Figure 6.4: Phosphine spectra with He/H_2 broadening at 1 bar pressure, using all the SAlTY transitions stronger than 10^{-30} cm^2/molecule for temperatures from 300 K to 2000 K. Figure provided through private correspondence.



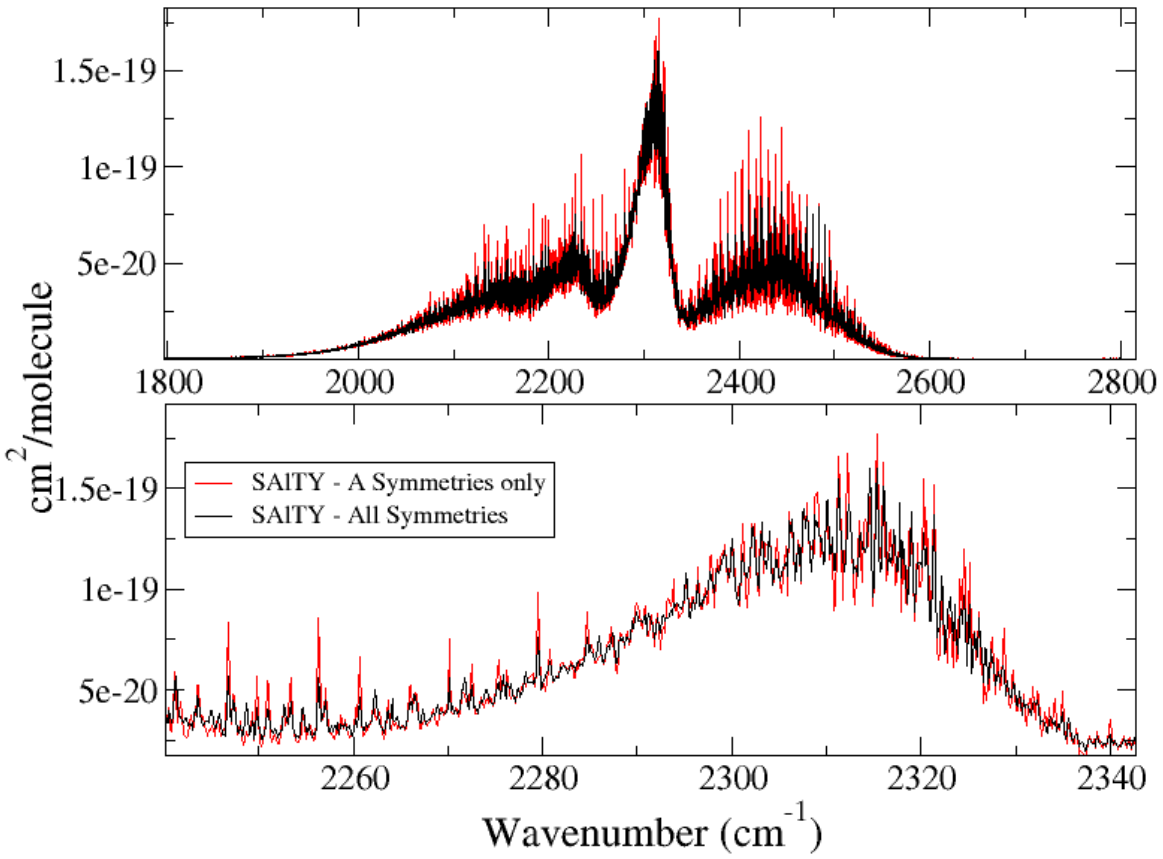
6.2 Phantom Symmetries

Spectral requirements for astrophysical modelling are usually low in resolution. An area of interest would be to study what effect lowering the accuracy of the line list calculations would have on spectra at different resolutions. In particular, at which resolution would a spectrum be inaccurate if only transitions of a particular symmetry were used to simulate it. The spectral structure of a molecule is largely determined by the electronic, rotational, and vibrational numbers, not the symmetry of the transitions. This is partially due to the fact that, as mentioned in chapter 4.1.3, for high Js, the actual energy values of a given J are very similar for all symmetries. It is therefore possible to roughly estimate the spectral contribution from transitions of an absent symmetry by carefully studying the transitions from the existing symmetries. The estimated contribution from the non-existing symmetry can then be manifested in the final spectrum as a ”phantom” opacity that roughly resembles its real equivalent. The study of this effect, which we have called Phantom Symmetries, is underway.

The Phantom Symmetries project will include an analysis of spectral shape accuracy costs when using only transitions of a selected symmetry, starting with phosphine as an example.

In phosphine, matrices for the E symmetry are roughly twice as large as those for A_1 and A_2 symmetries. The spectra due only to the A symmetries is similar to that due to the E symmetry and consequently, when its statistical weights are adjusted, resembles the total spectrum. This means that, at low resolutions, one symmetry's spectral contribution could be used to simulate another, allowing for a complete spectrum to be created using only a small selection of the calculated transitions. As can be seen from figure 6.5, the A symmetry spectra can replicate the complete SAlTY spectra reasonably well, even at relatively high resolution. Here the statistical weight given to the A_1 and A_2 symmetry transitions was doubled to account for the absence of E symmetry transitions.

Figure 6.5: Phosphine spectra at 1500 K, modelled using the complete SAlTY line list (in black) and using only the transitions between energy levels with A_1 and A_2 symmetry. Cross-sections with HWHM$= 0.1$ cm^{-1}.



In Chapters 3 and 5 it was discussed that the diagonalization of enormous matrices, and the calculation of billions of transitions, is an extremely costly process in terms of both time and power. With the phantom symmetries process, both could be drastically reduced, even if all symmetries in a molecule are equally difficult to simulate. The process becomes particularly useful, however, because of the size discrepancy between matrices of different symmetries. Within the same molecule, some symmetries have much smaller matrices than others, and as such are easier to diagonalize. The difficulty of diagonalization of an N by N matrix grows as N^3 in time and N^2 in memory, so the exclusion of large matrices from the simulation of a molecular spectra would drastically improve the calculation time and costs.

The Phantom Symmetries technique could be used to simulate spectra for molecules where little or no data exists, at a fraction of the time and computational power that full line lists would require. There are molecules for which matrices of particular symmetries are far too large for the current computational facilities to diagonlize. Until then, this technique could be used to create preliminary low resolution spectra which would be sufficiently accurate for many astronomical uses. Future work will explore the limitations of this approach for phosphine in terms of temperature, resolution and bin size, and then expand this investigation into other molecules.

6.3 Phosphine Tunnelling

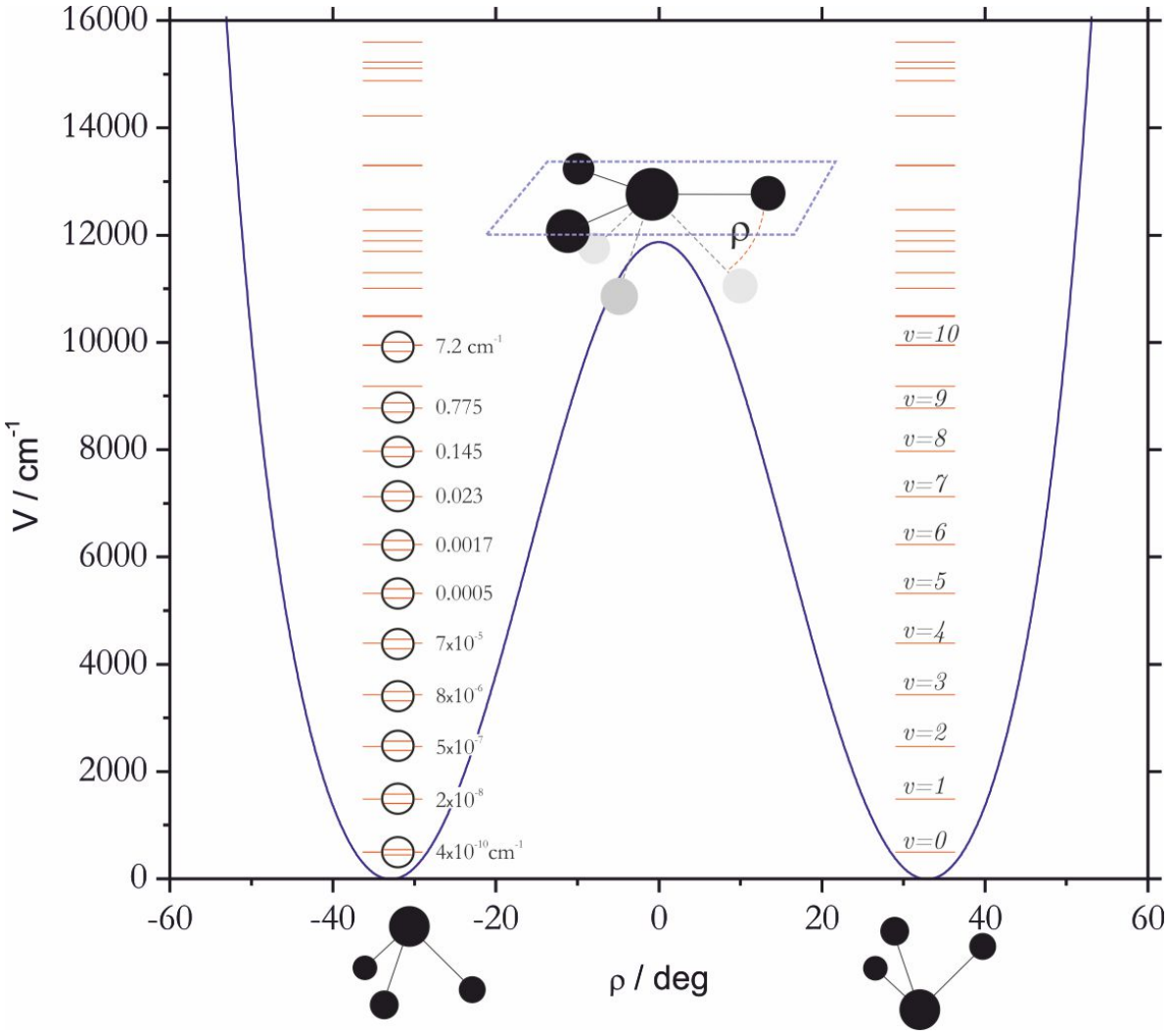
Tunnelling occurs when a pyramidal molecule goes through planarity and becomes inverted. This requires quantum tunnelling through an energy barrier and is most likely to occur if the barrier is sufficiently low and narrow. When a molecule undergoes quantum tunnelling, the superposition of the two states on either side of the energy barrier causes an energy level split and consequently a transition splitting, which can be visible in the molecule's spectrum. The inversion of the ammonia molecule through tunnelling was first detected by microwave spectroscopy in 1934 [36] and this tunnelling effect is predicted to also be found in phosphine [17] but, due to its much higher barrier (12 300 cm^{-1}) [148], is yet to be observed. The value of splitting in various vibrational states as well as the intensity of the inversion rotation and inversion-rovibrational lines can be computed by adapting the procedure used to simulate the phosphine spectrum to work with D_{3h}(M) symmetry, which is the permutation inversion group for ammonia, since it is much less rigid than phosphine.

An accurate description of the transitions that can be used to resolve the doublet splitting are part of ongoing work, but rough predictions for the position, size and intensity of transitions from split energy levels have been made, which are shown in Figure 6.6. The most promising region of possible detection is that of the symmetric bending band, $7\nu_2$ ($\approx$ 6900 cm^{-1}), which we predicted to have energy splittings of approximately 0.023 cm^{-1}. The most intense lines in the band are of the magnitude of 10^{-24} cm/molecule.

These splittings are of a reasonable magnitude to be resolved experimentally with the current level of spectroscopic precision, so it is expected the theoretical predictions of phosphine tunnelling shown here will be validated with experimental detection in the near future. Some observational spectroscopers, including Dr Livio Gianfrani from the Second University of Naples, have agreed to look further into this detection while convergence tests are made to ensure our values have sufficient precision and accuracy. Establishing convergence requires calculating energy levels with a very large basis set and a high polyad number, which is computationally expensive.

Work on the prediction of the phosphine splitting will be published in the near future, either from a purely theoretical perspective or, if possible, with the inclusion of a detection from experimental collaborators.

Figure 6.6: Preliminary splitting predictions for phosphine, showing the splitting for the ground state and the vibrational excitations up to $v = 10$ in the bending band ν_2.



6.4 Conclusion

The need for accurate and comprehensive molecular line lists is clear. Experimental processes are excellent sources of molecular data, but they are limited in relation to their completeness, energy level assignment, and high temperature spectra. The theoretical approach used in this thesis overcomes these limitations by generating a line list through *ab initio* quantum mechanical methods and state of the art high performance computing.

The ExoMol project aims to provide comprehensive line lists for every molecule relevant to the characterisation of the atmospheres of cool stars and exoplanets, and its team has spent the last four years developing high resolution theoretical spectra for dozens of molecules; the PH_3 molecule has been the author's focus. The final phosphine line list, SAlTY, is suitable for accurate simulations of spectra up to temperatures of 1500 K and it contains 16.8 billion transitions between 7.5 million energy levels. SAlTY's accuracy is validated by the excellent agreement at room temperatures with HITRAN, CDMS and other experimental sources and, like every other ExoMol molecule, it is freely available online in full.

The correct temperature dependence of line intensity requires an accurate understanding of the partition function, so high accuracy theoretical data has been used to calculate the temperature dependent partition function and related thermodynamical properties of phosphine and ammonia. The latter can be used in conjunction with the associated hot line list, BYTe, also created as part of the ExoMol project [183].

The development of the work presented here required sophisticated understanding of various aspects of quantum mechanics, molecular spectroscopy, planetary astrophysics, atmospheric modelling and computer science. Due to this multidisciplinary nature, it depended upon vast amounts of independent research and successful collaborative work with scientists from each field. It also required effective use of high performance computing and parallel processes. It ultimately led to the creation of an accurate and comprehensive line list for phosphine, SAlTY, and the associated thermodynamic data. This is a considerable improvement to the ability to model environments where the molecule is present, such as the atmospheres of cool stars and exoplanets.

Bibliography

[1] M. Agundez, J. Cernicharo, L. Decin, P. Encrenaz, and D. Teyssier. Confirmation of circumstellar phosphine. *Astrophys. J.*, 790:L27, 2014.

[2] A. Ainetschian, U. Haring, G. Spiegl, and W. Kreiner. The μ_2/μ_4 diad of PH_3. *J. Mol. Spectrosc.*, 181(1):99–107, 1997.

[3] A. F. Al-Refaie, J. Tennyson, and S. N. Yurchenko. GPU Acelerated INtensities (GAIN): a new method of computing Einstein A coefficients. *Comput. Phys. Commun.*, 2015. (to be submitted).

[4] A. F. Al-Refaie, S. N. Yurchenko, A. Yachmenev, and J. Tennyson. ExoMol line lists IX: A variationally computed line-list for hot formaldehyde. *Mon. Not. R. Astr. Soc.*, 2015.

[5] F. Allard, N. F. Allard, D. Homeier, J. Kielkopf, M. J. McCaughrean, and F. Spiegelman. K-H_2 quasi-molecular absorption detected in the T-dwarf epsilon Indi Ba. *Astron. Astrophys.*, 474(2):L21–L24, 2007.

[6] F. Allard, P. H. Hauschildt, D. R. Alexander, and S. Starrfield. Model atmospheres of very low mass stars and brown dwarfs. *Ann. Rev. Astron. Astrophys.*, 35(1):137–177, 1997.

[7] F. Allard, P. H. Hauschildt, I. Baraffe, and G. Chabrier. Synthetic spectra and mass determination of the brown dwarf Gliese 229B. *Astrophys. J. Lett.*, 465(2):L123, 1996.

[8] F. Allard, D. Homeier, B. Freytag, and C. Sharp. Atmospheres from very low-mass stars to extrasolar planets. *EAS Publications Series*, 57:3–43, 2012.

[9] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users'*

Guide. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999.

[10] J. Bailey and L. Kedziora-Chudczer. Modelling the spectra of planets, brown dwarfs and stars using vstar. *Mon. Not. R. Astr. Soc.*, 419(3):1913–1929, 2012.

[11] A. Baldacci, V. Malathy Devi, K. Narahari Rao, and G. Tarrago. Spectrum of phosphine at 4 to 5 μm: Analysis of ν_1 and ν_3 bands. *J. Mol. Spectrosc.*, 81:179–206, 1980.

[12] R. J. Barber, G. J. Harris, and J. Tennyson. Temperature dependent partition functions and equilibrium constant for HCN and HNC. *J. Chem. Phys.*, 117:11239–11243, 2002.

[13] R. J. Barber, S. Miller, N. Dello Russo, M. J. Mumma, J. Tennyson, and P. Guio. Water in the near IR spectrum of comet 8P/Tuttle. *Mon. Not. R. Astr. Soc.*, 398:1593–1600, 2009.

[14] R. J. Barber, J. Tennyson, G. J. Harris, and R. N. Tolchenov. A high accuracy computed water line list. *Mon. Not. R. Astr. Soc.*, 368:1087–1094, 2006.

[15] J. P. Beaulieu, G. Tinetti, D. Kipping, I. Ribas, R. J. Barber, J. Y.-K. Cho, I. Polichtchouk, J. Tennysson, S. N. Yurchenko, C. A. Griffith, I. Waldmann, S. Miller, S. Carey, O. Mousis, S. J. Fossey, and A. Aylward. Methane in the atmosphere of the transiting hot Neptune GJ436b? *Astrophys. J.*, 731:16, 2011.

[16] S. Belov, A. Burenin, L. Gershtein, A. Krupnov, V. Markov, A. Maslovsky, and S. Shapin. Rotational spectrum and spectroscopic constants of the phosphine molecule in the ground vibrational state. *J. Mol. Spectrosc.*, 86(1):184–192, 1981.

[17] S. P. Belov, A. V. Burenin, O. L. Polyansky, and S. M. Shapin. A new approach to the treatment of rotational spectra of molecules with small moments of inertia applied to the PH_3 molecule in the ground state. *J. Mol. Spectrosc.*, 90:579–589, 1981.

[18] M. Berblinger, C. Schlier, J. Tennyson, and S. Miller. Accurate specific molecular state densities by phase space integration: II comparison with quantum calculations on H_3^+ and HD_2^+. *J. Chem. Phys.*, 96:6842–6849, 1992.

[19] N. Biver and N. Biver. Cometary spectroscopy. *EAS Publications Series*, 47:165–188, 2011.

[20] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[21] J.-P. Bouanich, J. Salem, H. Aroui, J. Walrand, and G. Blanquet. H_2-broadening coefficients in the ν_2 and ν_4 bands of ph_3. *J. Quant. Spectrosc. Radiat. Transf.*, 84(2):195–205, 2004.

[22] D. Boyd and H. Longuet-Higgins. Coriolis interaction between vibration and rotation in symmetric top molecules. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 213(1112):55–73, 1952.

[23] R. A. Brooks and M. J. Mataric. Real robots, real learning problems. In *Robot learning*, pages 193–213. Springer, 1993.

[24] L. R. Brown, R. L. Sams, I. Kleiner, C. Cottaz, and L. Sagui. Line intensities of the phosphine dyad at 10 μm. *J. Mol. Spectrosc.*, 215:178–203, 2002.

[25] P. R. Bunker and P. Jensen. *Molecular Symmetry and Spectroscopy*. NRC Research Press, Ottawa, 2 edition, 1998.

[26] M. Burgdorf, G. Orton, T. Encrenaz, G. Davis, E. Lellouch, S. Sidher, and B. Swinyard. Far-infrared spectroscopy of the giant planets: measurements of ammonia and phosphine at jupiter and saturn and the continuum of neptune. *Advances in Space Research*, 34(11):2247–2250, 2004.

[27] J. L. Burgess and J. Burgess. Phosphine exposure from a methamphetamine laboratory investigation. *Clin. Tox.*, 39(2):165–168, 2001.

[28] R. A. H. Butler, L. Sagui, I. Kleiner, and L. R. Brown. The absorption spectrum of phosphine (ph_3) between 2.8 and 3.7 μm : Line positions, intensities, and assignments. *J. Mol. Spectrosc.*, 238:178–192, 2006.

[29] M. Carlotti, G. Di Lonardo, and A. Trombetti. K= 3 doubling in the v= 1 state of ph studied by laser-stark spectroscopy. *J. Chem. Phys.*, 78:1670–1673, 1983.

[30] B. E. Carlson, W. B. Rossow, and G. S. Orton. Cloud microphysics of the giant planets. *J. Atmos. Sci.*, 45:2066–2081, 1988.

[31] G. Cazzoli and C. Puzzarini. The lamb-dip spectrum of phosphine: The nuclear hyperfine structure due to hydrogen and phosphorus. *J. Mol. Spectrosc.*, 239:64–70, 2006.

[32] M. W. Chase and et al. *JANAF thermochemical tables: Journal of Physical and Chemical Reference Data*, volume 14. Washington, D.C : American Chemical Society ; New York : American Institute of Physics for the National Bureau of Standards., 1985.

[33] CHERIC. *www.cheric.org*, Korea Thermophysical Properties Data Bank.

[34] T. Y. Chesnokova, B. A. Voronin, A. D. Bykov, T. B. Zhuravleva, A. V. Kozodoev, A. A. Lugovskoy, and J. Tennyson. Calculation of solar radiation atmospheric absorption with different H_2O spectral line data banks. *J. Mol. Spectrosc.*, 256:41–44, 2009.

[35] F. Y. Chu and T. Oka. "forbidden" rotational spectra of phosphine and arsine. *J. Chem. Phys.*, 60:4612–4618, 1974.

[36] C. E. Cleeton and N. H. Williams. Electromagnetic waves of 1.1 cm wave-length and the absorption spectrum of ammonia. *Phys. Rev.*, 45(4):234, 1934.

[37] J. W. Cooley. *Math. Comp.*, 15:363–374, 1961.

[38] R. Courtin, D. Gautier, A. Marten, B. Bézard, and R. Hanel. The composition of saturn's atmosphere at northern temperate latitudes from Voyager IRIS spectra-NH_3, PH_3, C_2H_2, C_2H_6, CH_3D, CH_4, and the Saturnian D/H isotopic ratio. *Astrophys. J.*, 287:899–916, 1984.

[39] M. C. Cushing, M. S. Marley, D. Saumon, B. C. Kelly, W. D. Vacca, J. T. Rayner, R. S. Freedman, K. Lodders, and T. L. Roellig. Atmospheric parameters of field L and T dwarfs. *Astrophys. J.*, 678(2):1372, 2008.

[40] M. C. Cushing, J. T. Rayner, and W. D. Vacca. An infrared spectroscopic sequence of M, L, and T dwarfs. *Astrophys. J.*, 623(2):1115, 2005.

[41] P. B. Davies, R. M. Neumann, S. C. Wofsy, and W. Klemperer. Radio-frequency spectrum of phosphine (PH_3). *J. Chem. Phys.*, 55:3564–&, 1971.

[42] I. Dévai, L. Felföldy, I. Wittner, and S. Plósz. Detection of phosphine: new aspects of the phosphorus cycle in the hydrosphere. 1988.

[43] M. Dougherty, O. Grasset, E. Bunce, A. Coustenis, D. Titov, C. Erd, M. Blanc, A. Coates, A. Coradini, P. Drossart, et al. JUICE (JUpiter ICy moon Explorer): a European-led mission to the Jupiter system. In *EPSC-DPS Joint Meeting 2011*, volume 1, page 1343, 2011.

[44] M. J. Down, C. Hill, S. N. Yurchenko, J. Tennyson, L. R. Brown, and I. Kleiner. Re-analysis of ammonia spectra: Updating the HITRAN $^{14}NH_3$ database. *J. Quant. Spectrosc. Radiat. Transf.*, 130:260–272, 2013.

[45] P. Drossart, E. Lellouch, B. Bezard, J.-P. Maillard, and G. Tarrago. Jupiter: Evidence for a phosphine enhancement at high northern latitudes. *Icarus*, 83(1):248–253, 1990.

[46] T. H. Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.

[47] S. G. Edgington, S. K. Atreya, L. M. Trafton, J. J. Caldwell, R. F. Beebe, A. A. Simon, R. A. West, and C. Barnet. On the latitude variation of ammonia, acetylene, and phosphine altitude profiles on Jupiter from HST Faint Object Spectrograph observations. *Icarus*, 133:192–209, 1998.

[48] T. Encrenaz, T. Greathouse, P. Drossart, T. Fouchet, M. Janssen, S. Gulkis, G. Orton, L. Fletcher, and R. Giles. Monitoring jovian dynamics using maps of NH_3 and PH_3. *EPSC*, 9, 2014.

[49] B. Fegley Jr and K. Lodders. Atmospheric chemistry of the brown dwarf Gliese 229B: Thermochemical equilibrium predictions. *Astrophys. J.*, 472(1):L37, 1996.

[50] J. Fischer, R. R. Gamache, A. Goldman, L. S. Rothman, and A. Perrin. Total internal partition sums for molecular species in the 2000 edition of the HITRAN database. *J. Quant. Spectrosc. Radiat. Transf.*, 82:401–412, 2003.

[51] L. N. Fletcher, G. S. Orton, N. A. Teanby, and P. G. J. Irwin. Phosphine on Jupiter and Saturn from Cassini/CIRS. *Icarus*, 202:543–564, 2009.

[52] F. Fressin, G. Torres, D. Charbonneau, S. T. Bryson, J. Christiansen, C. D. Dressing, J. M. Jenkins, L. M. Walkowicz, and N. M. Batalha. The false positive rate of Kepler and the occurrence of planets. *Astrophys. J.*, 766(2):81, 2013.

[53] L. Fusina and M. Carlotti. The far-infrared spectrum and spectroscopic parameters of PH_3 in the ground state. *J. Mol. Spectrosc.*, 130(2):371–381, 1988.

[54] L. Fusina and G. Di Lonardo. The ν_2 and ν_4 bending fundamentals of phosphine (PH_3). *J. Mol. Spectrosc.*, 517:67–78, 2000.

[55] R. R. Gamache, S. Kennedy, R. Hawkins, and L. S. Rothman. Total internal partition sums for molecules in the terrestrial atmosphere. *J. Mol. Spectrosc.*, 517:407–425, 2000.

[56] G. Gassmann and D. Glindemann. Phosphane (ph3) in the biosphere. *Angewandte Chemie Int. Ed.*, 32(5):761–763, 1993.

[57] D. Glindemann, M. Edwards, and P. Kuschk. Phosphine gas in the upper troposphere. *Atm. Env.*, 37(18):2429–2433, 2003.

[58] A. Goldman, R. R. Gamache, A. Perrin, J.-M. Flaud, C. P. Rinsland, and L. S. Rothman. HITRAN partition functions and weighted transition-moments squared. *J. Quant. Spectrosc. Radiat. Transf.*, 66:455–486, 2000.

[59] R. S. Grammier. An overview of the Juno mission to Jupiter. 2006.

[60] N. N. Greenwood and A. Earnshaw. Chemistry of the elements. 1984.

[61] D. J. Griffiths and E. G. Harris. *Introduction to quantum mechanics*, volume 2. Prentice Hall New Jersey, 1995.

[62] M. U. Guide. The Mathworks. *Inc., Natick, MA*, 5:333, 1998.

[63] L. Haar. Thermodynamic properties of ammonia as an ideal gas. *J. Res. Nat. Bur. Standards A: Phys. Chem.*, 72:207–216, 1968.

[64] G. J. Harris, S. Viti, H. Y. Mussa, and J. Tennyson. Calculated high temperature partition function and related thermodynamic data for $H_2{}^{16}O$. *J. Chem. Phys.*, 109:7197–7204, 1998.

[65] S. G. He, J. J. Zheng, S. M. Hu, H. Lin, Y. Ding, X. H. Wang, and Q. S. Zhu. The stretching vibrational overtone spectra of PH_3: Local mode vibrational analysis, dipole moment surfaces from density functional theory and band intensities. *J. Chem. Phys.*, 114:7018–7027, 2001.

[66] C. Helling and S. Casewell. Atmospheres of brown dwarfs. *Astron. Astrophys. Rev.*, 22(1):1–45, 2014.

[67] P. Helminger and W. Gordy. Submillimeter-wave spectra of ammonia and phosphine. *Physical Review*, 188(1):100–107, 1969.

[68] D. A. Helms and W. Gordy. "forbidden" rotational spectra of symmetric-top molecules : PH_3 and PD_3. *J. Mol. Spectrosc.*, 66:206–218, 1977.

[69] A. J. Hewitt, N. Doss, N. F. Zobov, O. L. Polyansky, and J. Tennyson. Deuterated water: Partition functions and equilibrium constant. *Mon. Not. R. Astr. Soc.*, 356:1123–1126, 2005.

[70] C. Hill, I. E. Gordon, L. S. Rothman, and J. Tennyson. A new relational database structure and online interface for the HITRAN database. *J. Quant. Spectrosc. Radiat. Transf.*, 130:51–61, 2013.

[71] C. Hill, S. N. Yurchenko, and J. Tennyson. Temperature-dependent molecular absorption cross sections for exoplanets and other atmospheres. *Icarus*, 226:1673–1677, 2013.

[72] M. Hollis, M. Tessenyi, and G. Tinetti. Tau: A 1d radiative transfer code for transmission spectroscopy of extrasolar planet atmospheres. *Computer Physics Communications*, 184(10):2351–2361, 2013.

[73] R. Hu. Ammonia, water clouds and methane abundances of giant exoplanets and opportunities for super-earth exoplanets. *arXiv preprint arXiv:1412.7582*, 2014.

[74] A. W. Irwin. The partition functions of JANAF polyatomic molecules that significantly affect the stellar atmospheric equation of state. *Astrophys. J. Suppl.*, 74:145–160, 1988.

[75] N. Jacquinet-Husson, L. Crepeau, R. Armante, C. Boutammine, A. Chédin, N. Scott, C. Crevoisier, V. Capelle, C. Boone, N. Poulet-Crovisier, et al. The 2009 edition of the

GEISA spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.*, 112(15):2395–2445, 2011.

[76] L. Kaltenegger, W. Henning, and D. Sasselov. Detecting volcanism on extrasolar planets. *Astrophys. J.*, 140(5):1370, 2010.

[77] A. F. Krupnov, A. A. Melnikov, and V. A. Skvortsov. Comparison of dipole-moment of a PH_3 molecule in ground and excited-states. *Optika I Spektroskopiya*, 46:1012–1013, 1979.

[78] R. J. Kshirsagar. High-resolution FTIR spectroscopy of the $3\nu_2$ band of PH_3. *J. Mol. Spectrosc.*, 241:116–118, 2007.

[79] H. Kümmel. Origins of the coupled cluster method. *Theoretica chimica acta*, 80(2-3):81–89, 1991.

[80] V. Kunde, R. Hanel, W. Maguire, D. Gautier, J. Baluteau, A. Marten, A. Chedin, N. Husson, and N. Scott. The tropospheric gas composition of Jupiter's north equatorial belt/NH_3, PH_3, CH_3D, GeH_4, H_2O/and the Jovian D/H isotopic ratio. *Astrophys. J.*, 263:443–467, 1982.

[81] J. Kurzak, P. Luszczek, A. YarKhan, M. Faverge, J. Langou, H. Bouwmeester, J. Dongarra, J. J. Dongarra, M. Faverge, T. Herault, et al. Multithreading in the plasma library. *Multicore Computing: Algorithms, Architectures, and Applications*, page 119, 2013.

[82] J. H. Lacy, D. T. Jaffe, M. J. Richter, T. K. Greathouse, M. Bitner, P. Segura, W. Moller, T. R. Geballe, and K. Volk. TEXES on Gemini. In *SPIE Astronomical Telescopes+ Instrumentation*, pages 62694M–62694M. International Society for Optics and Photonics, 2006.

[83] M. Lambrechts, A. Johansen, and A. Morbidelli. Separating gas-giant and ice-giant planets by halting pebble accretion. *Astron. Astrophys.*, 572:A35, 2014.

[84] V. Laporta, R. Celiberto, and J. Tennyson. Resonant vibrational-excitation cross sections and rate constants for low-energy electron scattering by molecular oxygen. 22:025001, 2013.

[85] H. Lefebvre-Brion and R. W. Field. *The Spectra and Dynamics of Diatomic Molecules: Revised and Enlarged Edition*. Academic Press, 2004.

[86] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' Guide: Solution of Large-scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods (Software, Environments and Tools)*. Society for Industrial & Applied Mathematics, U.S., 1998. see http://www.caam.rice.edu/software/ARPACK/.

[87] E. G. Levels. Phosphine and eight metal phosphides acute exposure guideline levels. *National Academies Press (US)*, 2008.

[88] A. Levy, N. Lacome, and G. Tarrago. Hydrogen-and helium-broadening of phosphine lines. *J. Mol. Spectrosc.*, 157(1):172–181, 1993.

[89] G. F. Lindal, G. E. Wood, G. S. Levy, J. D. Anderson, D. N. Sweetnam, H. B. Hotz, B. J. Buckles, D. P. Holmes, P. E. Doms, V. R. Eshleman, et al. The atmosphere of Jupiter: An analysis of the Voyager radio occultation measurements. *J. Geophys. Res. A*, 86:8721–8727, 1981.

[90] E. Lindermeir and K. Beier. HITEMP derived spectral database for the prediction of jet engine exhaust infrared emission using a statistical band model. *J. Quant. Spectrosc. Radiat. Transf.*, 113(12):1575–1593, 2012.

[91] M. R. Line, J. Teske, B. Burningham, J. J. Fortney, and M. S. Marley. Uniform atmospheric retrieval analysis of ultracool dwarfs. i. characterizing benchmarks, Gl 570D and HD 3651B. *Astrophys. J.*, 807(2):183, 2015.

[92] K. Lodders. Revised Thermochemical Properties of Phosphinidene (PH), Phosphine (PH_3), Phosphorus Nitride (PN), and Magnesium Phosphate ($Mg_3P_2O_8$). *J. Phys. Chem. Ref. Data*, 28:1705–1712, 1999.

[93] L. Lodi and J. Tennyson. A linelist of allowed and forbidden rotational transitions for water. *J. Quant. Spectrosc. Radiat. Transf.*, 109:1219–1233, 2008.

[94] L. Lodi and J. Tennyson. Line lists for $H_2{}^{18}O$ and $H_2{}^{17}O$ based on empirically-adjusted line positions and ab initio intensities. *J. Quant. Spectrosc. Radiat. Transf.*, 113:850–858, 2012.

[95] L. Lodi, R. N. Tolchenov, J. Tennyson, A. E. Lynas-Gray, S. V. Shirin, N. F. Zobov, O. L. Polyansky, A. G. Császár, J. van Stralen, and L. Visscher. A new ab initio ground-state dipole moment surface for the water molecule. *J. Chem. Phys.*, 128:044304, 2008.

[96] C. Loomis and M. Strandberg. Microwave spectrum of phosphine, arsine, and stibine. *Phys. Rev.*, 81(5):798–807, 1951.

[97] P. W. Lucas, C. G. Tinney, B. Burningham, S. K. Leggett, D. J. Pinfield, R. Smart, H. R. A. Jones, F. Marocco, R. J. Barber, S. N. Yurchenko, J. Tennyson, M. Ishii, M. Tamura, A. C. Day-Jones, A. Adamson, F. Allard, and D. Homeier. The discovery of a very cool, very nearby brown dwarf in the Galactic plane. *Mon. Not. R. Astr. Soc.*, 408:L56–L60, 2010.

[98] A. E. Lynas-Gray, S. Miller, and J. Tennyson. Infra red transition intensities for water: a comparison of *ab initio* and fitted dipole moment surfaces. *J. Mol. Spectrosc.*, 169:458–467, 1995.

[99] E. Maciá. The role of phosphorus in chemical evolution. *Chem. Soc. Rev.*, 34(8):691–701, 2005.

[100] N. Madhusudhan, K. K. Lee, and O. Mousis. A possible carbon-rich interior in super-earth 55 cancri e. *Astrophys. J. Lett.*, 759(2):L40, 2012.

[101] N. Madhusudhan and S. Seager. A temperature and abundance retrieval method for exoplanet atmospheres. *Astrophys. J.*, 707(1):24, 2009.

[102] A. Maki, R. Sams, and W. Olson. Infrared determination of C for phosphine via perturbation-allowed δ k-l=$\pm$3 transitions in the 3ν band. *J. Chem. Phys.*, 58:4502, 1973.

[103] V. Malathy Devi, I. Kleiner, R. L. Sams, L. R. Brown, D. C. Benner, and L. N. Fletcher. Line positions and intensities of the phosphine pentad near 4.5 μm. *J. Mol. Spectrosc.*, 298:11–23, 2014.

[104] M. Marley, D. Saumon, T. Guillot, R. Freedman, W. Hubbard, A. Burrows, and J. Lunine. Atmospheric, evolutionary, and spectral models of the brown dwarf gliese 229 b. *Science*, 272(5270):1919–1921, 1996.

[105] M. S. Marley, A. S. Ackerman, J. N. Cuzzi, and D. Kitzmann. Clouds and hazes in exoplanet atmospheres. *Comp. Clim. Terr. Plan.*, 1:367–391, 2013.

[106] R. Marquardt, K. Sagui, J. Zheng, W. Thiel, D. Luckhaus, S. Yurchenko, F. Mariotti, and M. Quack. A global analytical potential energy surface for the electronic ground state of nh_3 from high level ab initio calculations. *J. Phys. Chem. A*, 2013.

[107] J. M. L. Martin, J.-P. François, and R. Gijbels. First principles computation of thermochemical properties beyond the harmonic approximation. I. Method and application to the water molecule and its isotopomers. *J. Chem. Phys.*, 96:7633, 1992.

[108] S. Miller, J. Tennyson, and B. T. Sutcliffe. Forbidden rotational and ro-vibrational tansitions in H_3^+: first principles calculations. *J. Mol. Spectrosc.*, 141:104–117, 1990.

[109] R. Moreno, A. Marten, and E. Lellouch. Search for PH_3 in the atmospheres of Uranus and Neptune at millimeter wavelength. In *AAS/Division for Planetary Sciences Meeting Abstracts 41*, volume 41, 2009.

[110] C. V. Morley, M. S. Marley, J. J. Fortney, R. Lupu, D. Saumon, T. Greene, and K. Lodders. Water clouds in Y dwarfs and exoplanets. *Astrophys. J.*, 787(1):78, 2014.

[111] J. I. Moses. Chemical kinetics on extrasolar planets. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2014):20130073, 2014.

[112] H. S. P. Müller, F. Schlöder, J. Stutzki, and G. Winnewisser. The Cologne Database for Molecular Spectroscopy, CDMS: a useful tool for astronomers and spectroscopists. *J. Mol. Struct.*, 742:215–227, 2005.

[113] D. F. Nava and L. J. Stief. Temperature study of oxygen atom+ phosphine reaction rate: kinetic measurements and planetary atmospheric implications. *J. Phys. Chem.*, 93(10):4044–4047, 1989.

[114] L. Neale, S. Miller, and J. Tennyson. Spectroscopic properties of the H_3^+ molecule: a new calculated linelist. *Astrophys. J.*, 464:516–520, 1996.

[115] L. Neale and J. Tennyson. A high temperature partition function for H_3^+. *Astrophys. J.*, 454:L169–L173, 1995.

[116] H.-U. Neue and H.-W. Scharpenseel. Gaseous products of the decomposition of organic matter in submerged soils. *Organi. Matt. Ric.*, pages 311–328, 1984.

[117] A. V. Nikitin, J. P. Champion, R. A. H. Butler, L. R. Brown, and I. Kleiner. Global modeling of the lower three polyads of PH_3: Preliminary results. *J. Mol. Spectrosc.*, 256:4–16, 2009.

[118] A. V. Nikitin, F. Holka, V. G. Tyuterev, and J. Fremont. Vibration energy levels of the PH_3, PH_2D, and PHD_2 molecules calculated from high order potential energy surface. *J. Chem. Phys.*, 130:244312, 2009.

[119] A. V. Nikitin, M. Rey, and V. G. Tyuterev. High order dipole moment surfaces of PH_3 and ab initio intensity predictions in the octad range. *J. Mol. Spectrosc.*, pages –, 2014. (in press).

[120] C. A. Nixon, N. A. Teanby, P. G. Irwin, and S. M. Hörst. Upper limits for PH_3 and H_2S in Titan's atmosphere from cassini cirs. *Icarus*, 224(1):253–256, 2013.

[121] J. Noga and R. J. Bartlett. The full CCSDT model for molecular electronic structure. *J. Chem. Phys.*, 86(12):7041–7050, 1987.

[122] K. S. Noll. Spectroscopy of outer solar system atmospheres from 2.5 to 7.0 microns. In *Astronomical Infrared Spectroscopy: Future Observational Directions*, volume 41, page 29, 1993.

[123] K. S. Noll and M. S. Marley. Detectability of CO, PH_3, AsH_3, and GeH_4 in the atmosphere of GL 229B. In *Planets Beyond the Solar System and the Next Generation of Space Missions*, volume 119, page 115, 1997.

[124] B. Numerov. *Mon. Not. R. Astron. Soc.*, 84:592–xxx, 1924.

[125] B. Oppenheimer, C. Baranec, C. Beichman, D. Brenner, R. Burruss, E. Cady, J. Crepp, R. Dekany, R. Fergus, D. Hale, et al. Reconnaissance of the HR 8799 exosolar system. I. Near-infrared spectroscopy. *Astrophys. J.*, 768(1):24, 2013.

[126] B. Oppenheimer, S. Kulkarni, K. Matthews, and M. Van Kerkwijk. The spectrum of the brown dwarf gliese 229b. *Astrophys. J.*, 502(2):932, 1998.

[127] R. I. Ovsyannikov, W. Thiel, S. N. Yurchenko, M. Carvajal, and P. Jensen. PH_3 revisited: Theoretical transition moments for the vibrational transitions below. *J. Mol. Spectrosc.*, 252:121–128, 2008.

[128] R. I. Ovsyannikov, W. Thiel, S. N. Yurchenko, M. Carvajal, and P. Jensen. Vibrational energies of PH_3 calculated variationally at the complete basis set limit. *J. Chem. Phys.*, 129:044309, 2008.

[129] T. Owen and T. Encrenaz. Element abundances and isotope ratios in the giant planets and Titan. In *Solar System History from Isotopic Signatures of Volatile Elements*, pages 121–138. Springer, 2003.

[130] J. C. Pearson, H. S. P. Müller, H. M. Pickett, E. A. Cohen, and B. J. Drouin. Introduction to submillimeter, millimeter and microwave spectral line catalog. *J. Quant. Spectrosc. Radiat. Transf.*, 111(11):1614–1616, 2010.

[131] M. Perryman. *The exoplanet handbook*. Cambridge University Press, 2011.

[132] H. M. Pickett, R. L. Poynter, and E. A. Cohen. Pressure broadening of phosphine by hydrogen and helium. *J. Quant. Spectrosc. Radiat. Transf.*, 26:197–198, 1981.

[133] H. M. Pickett, R. L. Poynter, E. A. Cohen, M. L. Delitsky, J. C. Pearson, and H. S. P. Müller. Submillimeter, millimeter, and microwave spectral line catalog. *J. Quant. Spectrosc. Radiat. Transf.*, 60:883–890, 1998.

[134] O. L. Polyansky, I. N. Kozin, P. Maĺyszek, J. Koput, J. Tennyson, and S. N. Yurchenko. Variational calculation of highly excited rovibrational energy levels of H_2O_2. *J. Phys. Chem. A*, 117:73677377, 2013.

[135] R. G. Prinn and J. S. Lewis. Phosphine on Jupiter and implications for the Great Red Spot. *Science*, 190:274–276, 1975.

[136] M. Quack. *Fundamental symmetries and symmetry violations from high resolution spectroscopy*. Wiley Online Library, 2011.

[137] J. T. Rayner, M. C. Cushing, and W. D. Vacca. The infrared telescope facility (IRTF) spectral library: cool stars. *Astrophys. J. Suppl.*, 185(2):289, 2009.

[138] E. L. Rice, T. Barman, I. S. Mclean, L. Prato, and J. D. Kirkpatrick. Physical properties of young brown dwarfs and very low mass stars inferred from high-resolution model spectra. *Astrophys. J. Suppl.*, 186(1):63, 2010.

[139] L. S. Rothman, I. E. Gordon, Y. Babikov, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, L. Bizzocchi, V. Boudon, L. R. Brown, A. Campargue, K. Chance, E. A. Cohen, L. H. Coudert, V. M. Devi, B. J. Drouin, A. Fayt, J.-M. Flaud, R. R. Gamache, J. J. Harrison, J.-M. Hartmann, C. Hill, J. T. Hodges, D. Jacquemart, A. Jolly, J. Lamouroux, R. J. Le Roy, G. Li, D. A. Long, O. M. Lyulin, C. J. Mackie, S. T. Massie, S. Mikhailenko, H. S. P. Müller, O. V. Naumenko, A. V. Nikitin, J. Orphal, V. Perevalov, A. Perrin, E. R. Polovtseva, C. Richard, M. A. H. Smith, E. Starikova, K. Sung, S. Tashkun, J. Tennyson, G. C. Toon, V. G. Tyuterev, and G. Wagner. The *HITRAN* 2012 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.*, 130:4 – 50, 2013.

[140] L. S. Rothman, I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J. P. Champion, K. Chance, L. H. Coudert, V. Dana, V. M. Devi, S. Fally, J. M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, I. Kleiner, N. Lacome, W. J. Lafferty, J. Y. Mandin, S. T. Massie, S. N. Mikhailenko, C. E. Miller, N. Moazzen-Ahmadi, O. V. Naumenko, A. V. Nikitin, J. Orphal, V. I. Perevalov, A. Perrin, A. Predoi-Cross, C. P. Rinsland, M. Rotger, M. Simeckova, M. A. H. Smith, K. Sung, S. A. Tashkun, J. Tennyson, R. A. Toth, A. C. Vandaele, and J. Vander Auwera. The *HITRAN* 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.*, 110:533–572, 2009.

[141] L. S. Rothman, I. E. Gordon, R. J. Barber, H. Dothe, R. R. Gamache, A. Goldman, V. I. Perevalov, S. A. Tashkun, and J. Tennyson. HITEMP, the High-Temperature Molecular Spectroscopic Database. *J. Quant. Spectrosc. Radiat. Transf.*, 111:2139–2150, 2010.

[142] N. D. Russo, B. Bonev, M. DiSanti, M. Mumma, E. Gibb, K. Magee-Sauer, R. Barber, and J. Tennyson. Water production rates, rotational temperatures, and spin temperatures in comets C/1999 H1 (Lee), C/1999 S4, and C/2001 A2. *Astrophys. J.*, 621(1):537, 2005.

[143] J. Salem, H. Aroui, J. P. Bouanich, J. Walrand, and G. Blanquet. Collisional broadening and line intensities in the ν_2 and ν_4 bands of ph_3. *J. Mol. Spectrosc.*, 225(2):174–181, 2004.

[144] J. Salem, G. Blanquet, M. Lepère, and H. Aroui. H_2 line mixing coefficients in the ν_2 and ν_4 bands of PH_3. *J. Mol. Spectrosc.*, 297:58–61, 2014.

[145] D. Saumon, M. Marley, M. Cushing, S. Leggett, T. Roellig, K. Lodders, and R. Freedman. Ammonia as a tracer of chemical equilibrium in the T7. 5 dwarf Gliese 570D. *Astrophys. J.*, 647(1):552, 2006.

[146] D. Saumon, M. S. Marley, M. Abel, L. Frommhold, and R. S. Freedman. New H_2 collision-induced absorption and NH_3 opacity and the spectra of the coolest brown dwarfs. *Astrophys. J.*, 750:74, 2012.

[147] D. Schleicher, D. Galli, F. Palla, M. Camenzind, R. Klessen, M. Bartelmann, and S. Glover. Effects of primordial chemistry on the cosmic microwave background. *Astron. Astrophys.*, 490(2):521–535, 2008.

[148] P. Schwerdtfeger, L. Laakkonen, and P. Pyykko. Trends in inversion barriers. I. Group-15 hydrides. *J. Chem. Phys.*, 96:6807–6819, 1992.

[149] S. Seager, M. Kuchner, C. Hier-Majumder, and B. Militzer. Mass-radius relationships for solid exoplanets. *Astrophys. J.*, 669(2):1279, 2007.

[150] S. W. Sharpe, T. J. Johnson, R. L. Sams, P. M. Chu, G. C. Rhoderick, and P. A. Johnson. Gas-phase databases for quantitative infrared spectroscopy. *Applied Spec.*, 58(12):1452–1461, 2004.

[151] A. P. Showman, J. J. Fortney, Y. Lian, M. S. Marley, R. S. Freedman, H. A. Knutson, and D. Charbonneau. Atmospheric circulation of hot Jupiters: Coupled radiative-dynamical general circulation model simulations of HD 189733b and HD 209458b. *Astrophys. J.*, 699(1):564, 2009.

[152] M. Šimečková, D. Jacquemart, L. S. Rothman, R. R. Gamache, and A. Goldman. Einstein A coefficients and statistical weights for molecular absorption transitions in the HITRAN database. *J. Quant. Spectrosc. Radiat. Transf.*, 98:130–155, 2006.

[153] I. Snellen, S. Albrecht, E. de Mooij, and R. Le Poole. Ground-based detection of sodium in the transmission spectrum of exoplanet HD 209458b. *Astron. Astrophys.*, 487(1):357–362, 2008.

[154] I. Snellen, R. de Kok, J. Birkby, B. Brandl, M. Brogi, C. Keller, M. Kenworthy, H. Schwarz, and R. Stuik. Combining high-dispersion spectroscopy with high contrast imaging: Probing rocky planets around our nearest neighbors. *Astron. Astrophys.*, 576:A59, 2015.

[155] C. Suarez. Line intensity measurement for the ν_1 and ν_3 bands of phosphine. *Spectroscopy letters*, 35(6):757–767, 2002.

[156] M. Swain, G. Vasisht, G. Tinetti, J. Bouwman, P. Chen, Y. Yung, D. Deming, and P. Deroo. Molecular signatures in the near-infrared dayside spectrum of HD 189733b. *Astrophys. J. Lett.*, 690(2):L114, 2009.

[157] M. R. Swain, G. Vasisht, and G. Tinetti. The presence of methane in the atmosphere of an extrasolar planet. *Nature*, 452(7185):329–331, 2008.

[158] G. Tarrago, M. Dang-Nhu, and A. Goldman. Analysis of phosphine absorption in the region 9-10 μm and high-resolution line-by-line simulation of the ν_2 and ν_4 bands. *J. Mol. Spectrosc.*, 88:311–322, 1981.

[159] G. Tarrago, N. Lacome, A. Levy, G. Guelachvili, B. Bezard, and P. Drossart. Phosphine Spectrum at 4-5 μm : Analysis and Line-by-Line Simulation of $2\nu_2$, $\nu_2 + \nu_4$, ν_1 and ν_3 Bands. *J. Mol. Spectrosc.*, 154:30–42, 1992.

[160] J. Tennyson. *Astronomical spectroscopy: an introduction to the atomic and molecular physics of astronomical spectra*. World Scientific, 2010.

[161] J. Tennyson. *High Accuracy Rotation-vibration Calculations on Small Molecules*, pages 551–572. John Wiley & Sons, Chichester, UK, 2011.

[162] J. Tennyson. Vibration-rotation transition intensities from first principles. *J. Mol. Spectrosc.*, 298:1–6, 2014.

[163] J. Tennyson, P. F. Bernath, A. Campargue, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, D. Lisak, O. V. Naumenko, L. S. Rothman, H. Tran, N. F. Zobov,

J. Buldyreva, C. D. Boone, M. D. De Vizia, L. Gianfrani, J.-M. Hartmann, R. McPheat, J. Murray, N. H. Ngo, O. L. Polyansky, and D. Weidmann. Recommended isolated-line profile for representing high-resolution spectroscopic transitions (IUPAC Technical Report). *Pure Appl. Chem.*, 86:1931–1943, 2014.

[164] J. Tennyson, C. Hill, and S. N. Yurchenko. Data structures for ExoMol: Molecular line lists for exoplanet and other atmospheres. In *6^{th} international conference on atomic and molecular data and their applications ICAMDATA-2012*, volume 1545 of *AIP Conference Proceedings*, pages 186–195. AIP, New York, 2013.

[165] J. Tennyson and S. N. Yurchenko. ExoMol: molecular line lists for exoplanet and other atmospheres. *Mon. Not. R. Astr. Soc.*, 425:21–33, 2012.

[166] G. Tinetti, P. Deroo, M. Swain, C. Griffith, G. Vasisht, L. Brown, C. Burke, and P. McCullough. Probing the terminator region atmosphere of the hot-jupiter XO-1b with transmission spectroscopy. *Astrophys. J. Lett.*, 712(2):L139, 2010.

[167] G. Tinetti, A. Vidal-Madjar, M.-C. Liang, J.-P. Beaulieu, Y. Yung, S. Carey, R. J. Barber, J. Tennyson, I. Ribas, N. Allard, G. E. Ballester, D. K. Sing, and F. Selsis. Water vapour in the atmosphere of a transiting extrasolar planet. *Nature*, 448:169–171, 2007.

[168] T. Tsuji, K. Ohnaka, W. Aoki, and T. Nakajima. Evolution of dusty photospheres through red to brown dwarfs: how dust forms in very low mass objects. *Astron. Astrophys.*, 308:L29–L32, 1996.

[169] O. N. Ulenikov, E. S. Bekhtereva, V. A. Kozinskaia, J. J. Zheng, S. G. He, S. M. Hu, Q. S. Zhu, C. Leroy, and L. Pluchart. On the study of resonance interactions and splittings in the PH_3 molecule: ν_1, ν_3, ν_2+ν_4, $2\nu_4$ bands. *J. Mol. Spectrosc.*, 215:295–308, 2002.

[170] O. N. Ulenikov, E. S. Bekhtereva, V. A. Kozinskaia, J. J. Zheng, S. G. He, S. M. Hu, Q. S. Zhu, C. Leroy, and L. Pluchart. High-resolution spectrum of the ν_1+ ν_4 (E), ν_3+ ν_4 (E), ν_3+ ν_4 (A1), and ν_3 + ν_4 (A2) bands of the PH_3 molecule: assignments and preliminary analysis. *J. Quant. Spectrosc. Radiat. Transf.*, 83:599–618, 2004.

[171] D. S. Underwood, J. Tennyson, and S. N. Yurchenko. An ab initio variationally computed room-temperature line list for SO_3. *Phys. Chem. Chem. Phys.*, 15:10118–10125, 2013.

[172] M. Vidler and J. Tennyson. Accurate partition function and thermodynamic data for water. *J. Chem. Phys.*, 113:9766–9771, 2000.

[173] C. Visscher, K. Lodders, and B. Fegley Jr. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. II. Sulfur and Phosphorus. *Astrophys. J.*, 648(2):1181, 2006.

[174] D. S. Visualizer. Accelrys software inc. *Discovery Studio Visualizer*, 2, 2005.

[175] I. Waldmann et al. Of "cocktail parties" and exoplanets. *Astrophys. J.*, 747(1):12, 2012.

[176] D. Wang, Q. Shi, and Q.-S. Zhu. An ab initio quartic force field of PH_3. *J. Chem. Phys.*, 112:21, 2000.

[177] L. Wang, P. Chen, G. Cheng, Y. Ding, and S. Hu. Absorption line intensities of phosphine in the regions 1 950-2 480 cm^{-1} and 3 280-3 580 cm^{-1} studied by fourier-transform spectroscopy. *Guang pu xue yu guang pu fen xi*, 25(8):1221, 2005.

[178] M. Werner, T. Roellig, F. Low, G. Rieke, M. Rieke, W. Hoffmann, E. Young, J. Houck, B. Brandl, G. Fazio, et al. The Spitzer space telescope mission. *Astrophys. J. Suppl.*, 154(1):1, 2004.

[179] R. Wilson, F. H. Lovejoy, R. J. Jaeger, and P. L. Landrigan. Acute phosphine poisoning aboard a grain freighter: epidemiologic, clinical, and pathological findings. *JAMA*, 244(2):148–150, 1980.

[180] M. Windholz, S. Budavari, L. Y. Stroumtsos, M. N. Fertig, et al. *The Merck index. An encyclopedia of chemicals and drugs.* Number 9th edition. Merck & Co., 1976.

[181] A. Yachmenev, S. N. Yurchenko, P. Jensen, and W. Thiel. A new "spectroscopic" potential energy surface for formaldehyde in its ground electronic state. *J. Chem. Phys.*, 134:244307, 2011.

[182] A. Yachmenev, S. N. Yurchenko, I. Paidarova, P. Jensen, W. Thiel, and S. P. A. Sauer. Thermal averaging of the indirect nuclear spin-spin coupling constants of ammonia: The importance of the large amplitude inversion mode. *J. Chem. Phys.*, 132:114305, 2010.

[183] S. N. Yurchenko, R. J. Barber, and J. Tennyson. A variationally computed hot line list for NH_3. *Mon. Not. R. Astr. Soc.*, 413:1828–1834, 2011.

[184] S. N. Yurchenko, R. J. Barber, J. Tennyson, W. Thiel, and P. Jensen. Towards efficient refinement of molecular potential energy surfaces: Ammonia as a case study. *J. Mol. Spectrosc.*, 268:123–129, 2011.

[185] S. N. Yurchenko, R. J. Barber, A. Yachmenev, W. Thiel, P. Jensen, and J. Tennyson. A variationally computed T=300 K line list for NH_3. *J. Phys. Chem. A*, 113:11845–11855, 2009.

[186] S. N. Yurchenko, M. Carvajal, P. Jensen, F. Herregodts, and T. R. Huet. Potential parameters of PH_3 obtained by simultaneous fitting of ab initio data and experimental vibrational band origins. *Chem. Phys.*, 290:59–67, 2003.

[187] S. N. Yurchenko, M. Carvajal, W. Thiel, and P. Jensen. Ab initio dipole moment and theoretical rovibrational intensities in the electronic ground state of PH_3. *J. Mol. Spectrosc.*, 239:71–87, 2006.

[188] S. N. Yurchenko, M. Carvajal, A. Yachmenev, W. Thiel, and P. Jensen. A theoretical-spectroscopy, ab initio-based study of the electronic ground state of (SbH_3)-Sb-121. *J. Quant. Spectrosc. Radiat. Transf.*, 111:2279–2290, 2010.

[189] S. N. Yurchenko, M. Carvajalz, P. Jensen, H. Lin, J. J. Zheng, and W. Thiel. Rotation-vibration motion of pyramidal XY3 molecules described in the Eckart frame: Theory and application to NH_3. *Mol. Phys.*, 103:359–378, 2005.

[190] S. N. Yurchenko and J. Tennyson. ExoMol line lists IV: The rotation-vibration spectrum of methane up to 1500 K. *Mon. Not. R. Astr. Soc.*, 440:1649–1661, 2014.

[191] S. N. Yurchenko, J. Tennyson, J. Bailey, M. D. J. Hollis, and G. Tinetti. Spectrum of hot methane in astronomical objects using a comprehensive computed line list. *Proc. Nat. Acad. Sci.*, 111:9379–9383, 2014.

[192] S. N. Yurchenko, W. Thiel, M. Carvajal, H. Lin, and P. Jensen. Rotation-vibration motion of pyramidal XY3 molecules described in the Eckart frame: The calculation of intensities with application to NH_3. *Adv. Quant. Chem.*, 48:209–238, 2005.

[193] S. N. Yurchenko, W. Thiel, and P. Jensen. Theoretical ROVibrational Energies (TROVE): A robust numerical approach to the calculation of rovibrational energies for polyatomic molecules. *J. Mol. Spectrosc.*, 245:126–140, 2007.

[194] S. N. Yurchenko, W. Thiel, S. Patchkovskii, and P. Jensen. Theoretical evidence for the formation of rotational energy level clusters in the vibrational ground state of PH_3. *Phys. Chem. Chem. Phys.*, 7:573–582, 2005.

[195] P. Zhou and Y. Chen. Molecular environment and an X-ray spectroscopy of supernova remnant Kesteven 78. *Astrophys. J.*, 743(1):4, 2011.

[196] R. Zhu, L. Sun, D. Kong, J. Geng, N. Wang, Q. Wang, and X. Wang. Matrix-bound phosphine in antarctic biosphere. *Chemosphere*, 64(8):1429–1435, 2006.